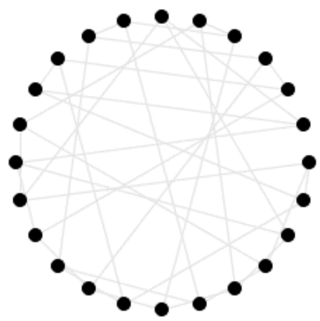
NYU STERN
NEW YORK UNIVERSITY · LEONARD N. STERN SCHOOL OF BUSINESS
CENTER FOR DIGITAL ECONOMY RESEARCH

Modeling Complex Networks For (Electronic) Commerce

Foster Provost, Arun Sundararajan
ACM EC'07 – June 12, 2007



Why do networks matter in commerce?

- What are examples of "large sets of irregularly connected entities" we observe as a consequence of (electronic) commerce?

(intentionally blank)

Why are these "networked" data valuable?

Why do networks matter in commerce?

- What are examples of “large sets of irregularly connected entities” that affect outcomes in (electronic) commerce and which we <u>do not</u> observe ?

(intentionally blank)

What explains the formation and structure of these “underlying” networks?

A very basic framework

- There are underlying networks that affect outcomes in electronic commerce.
  - Manageable and useful abstractions of these networks which are informed by theories from the social sciences can lead to better theories that are related to electronic commerce.
- There are empirical networks generated as a by-product of electronic commerce which can
  - Describe outcomes of electronic commerce;
  - Be used to predict future outcomes, and
  - Influence underlying networks.
- Modeling these empirical networks in a rigorous way can be informed by useful abstractions of the underlying networks that generate them.

Agenda for this tutorial

- Abstracting networks towards better theory.
- Modeling for prediction using networked data.
- Modeling for explanation using networked data.

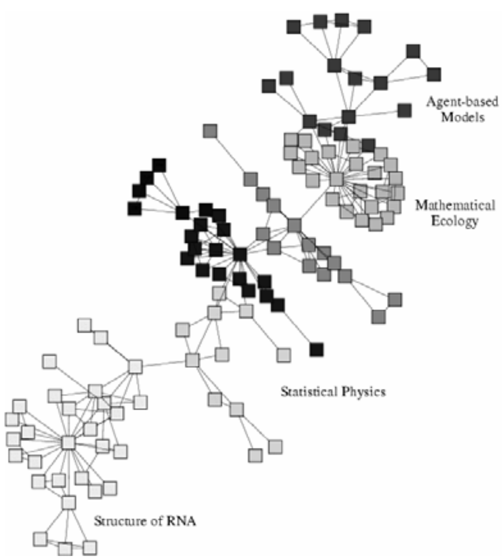
(1) Abstracting networks to theorize

Abstracting networks to theorize

Goals of this part of the tutorial

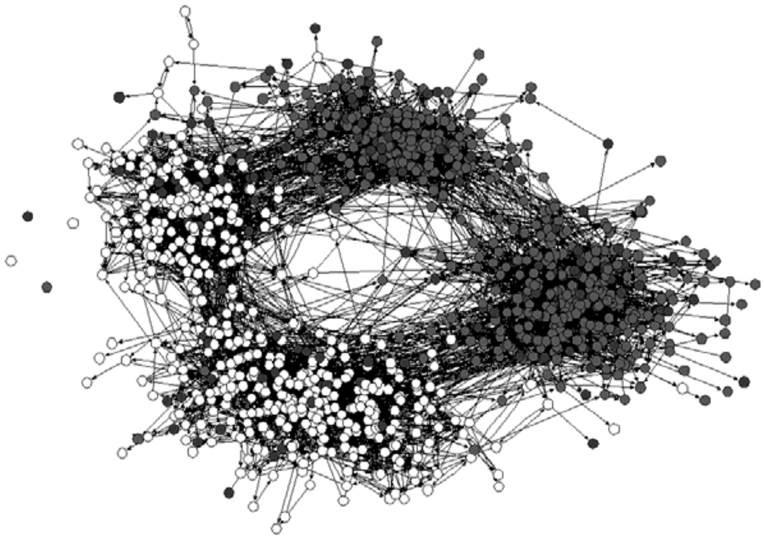
- A basic understanding of the diversity of "complex" networks in business, society and nature
- A basic understanding of some properties of these networks that are useful.
- A basic understanding of the manageable mathematical abstractions of these networks, and the connection between these abstractions and the properties described above.

Examples of networks



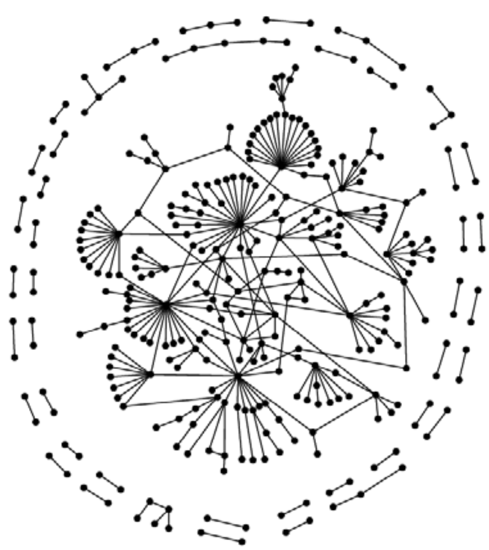
Interdisciplinary collaboration network at the Santa Fe Institute

Examples of networks



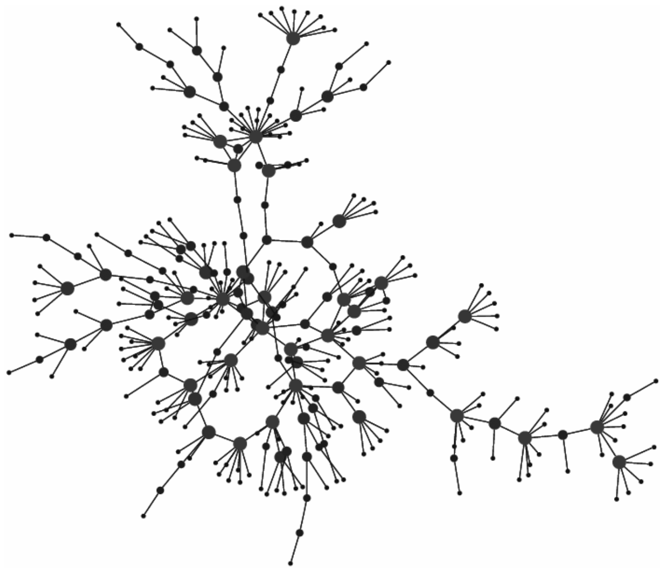
High-school friendship network

Examples of networks



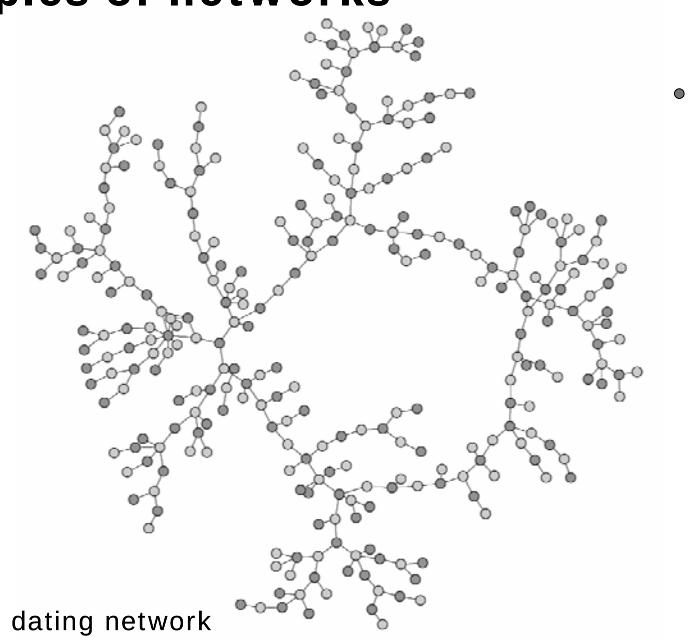
Yeast network

Examples of networks



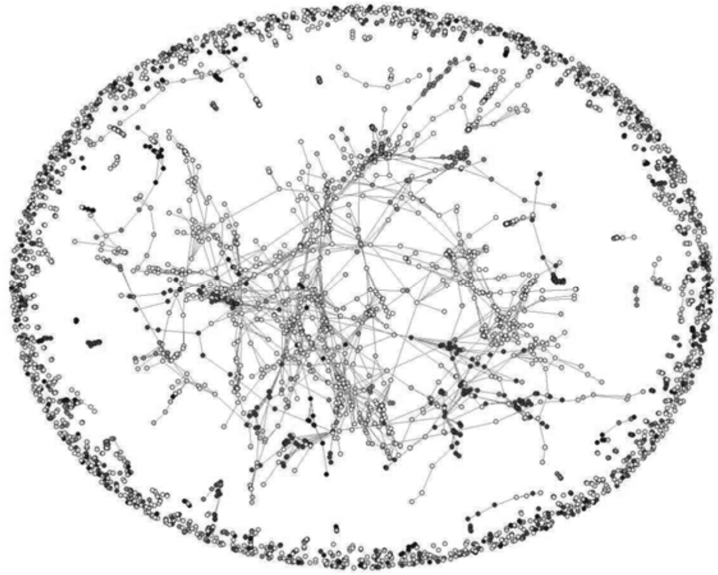
Sexual contact network

Examples of networks



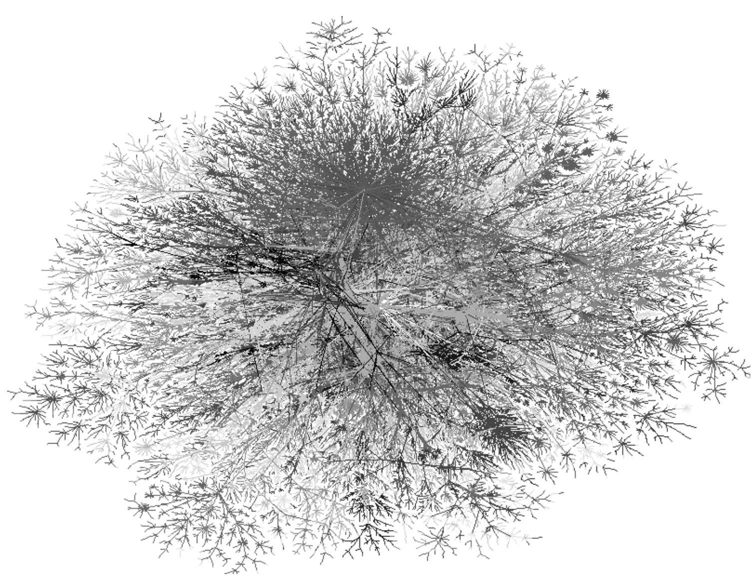
High-school dating network

Examples of networks



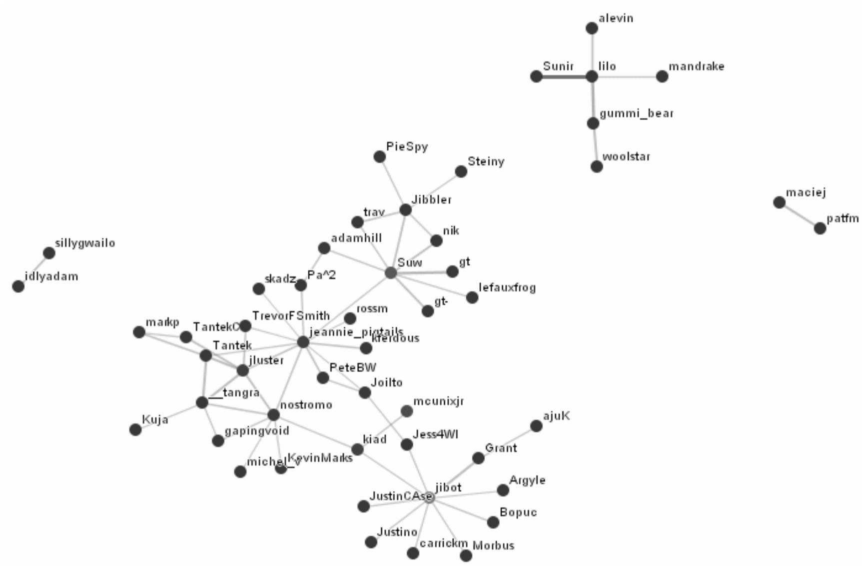
Machine Learning Papers

Examples of networks



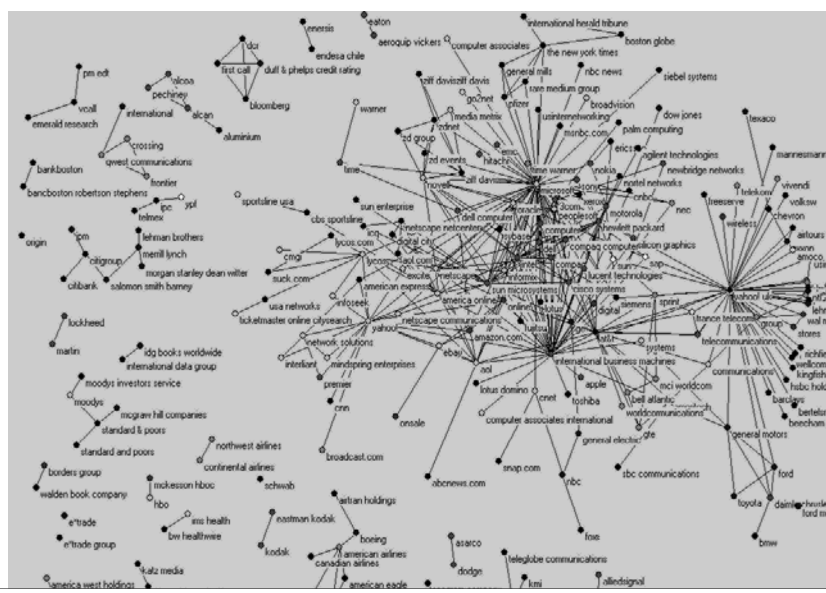
The Web, circa 1998

Examples of networks



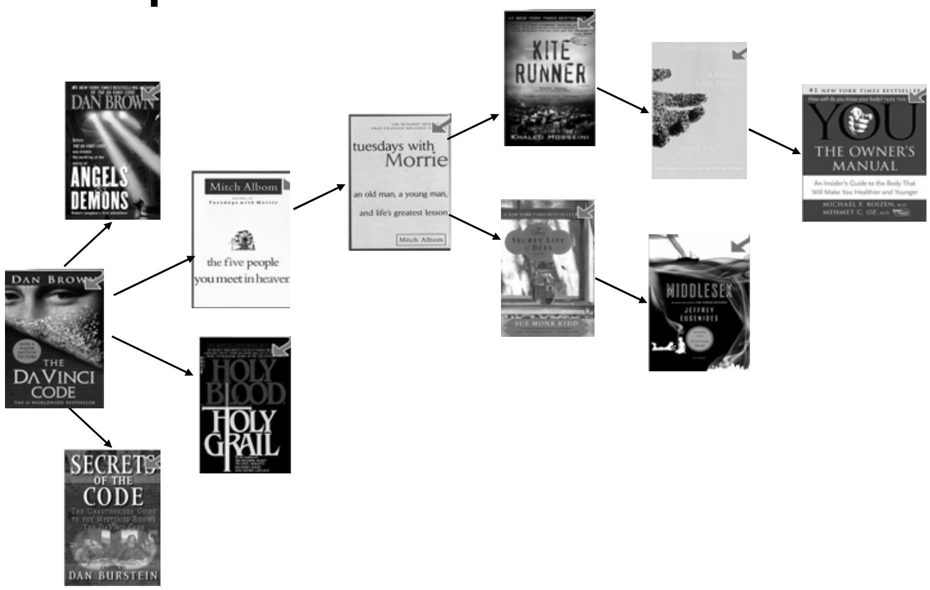
IRC channel

Examples of networks



(Companies linked by news stories)

Examples of networks



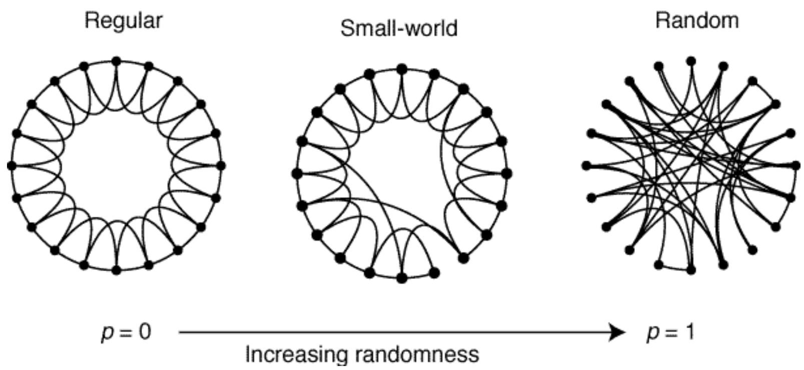
Books linked by co-purchases (partial...)

Overview of networks

- Some basic terminology
  - Graph
  - Node, edge
  - Directed/undirected
  - Degree (degree distribution)
  - Component

Random graphs

- Analogous to random variables
- Poisson (Erdos-Renyi) random graph: $q(x) = \binom{n}{k} p^k (1-p)^{n-k}$
- Generalized random graphs
- Models of small-world graphs



Random graphs, more importantly…

- Conceptual construct for modeling networks
- Simplest abstraction: a graph is drawn from a set of possible graphs according to some distribution
- More useful but less precise abstraction
  - The distributions associated with the **properties** of the graphs that are eventually drawn.
  - So, what exactly is a network **property**?

Network properties

- Degree distribution
  - Extent of and variation in "local connectedness" across nodes
- PageRank (and related measures)
  - Extent of and variation in "centrality" across nodes
- Clustering
  - Extent of and variation in "shared connectedness" across nodes
- Average distance (diameter)
  - Extent of and variation in distance between nodes
- Assortative mixing/Homophily
  - Extent of and variation in "within-class connectedness" across nodes
- Distribution of components, degree correlation, community structure,…

Random graphs

- Conceptual construct for modeling networks
- Simplest abstraction: a graph is drawn from a set of possible graphs according to some distribution
- Simplest less precise abstraction:
  - Each draw is described in terms of a degree distribution

    $q(x)$: fraction of nodes with degree x
  - Need independence assumptions, a construction process
  - Power-law networks: $q(x) = x^{-\alpha}$

Examples

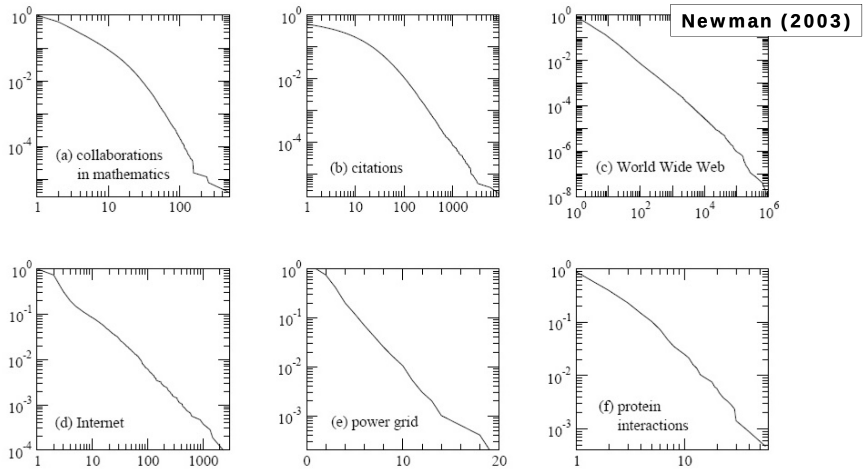
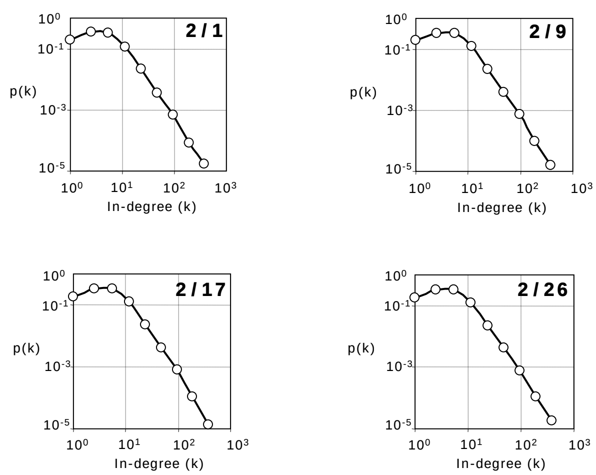


FIG. 6 Cumulative degree distributions for six different networks. The horizontal axis for each panel is vertex degree k (or in-degree for the citation and Web networks, which are directed) and the vertical axis is the cumulative probability distribution of degrees, i.e., the fraction of vertices that have degree greater than or equal to k. The networks shown are: (a) the collaboration network of mathematicians [182]; (b) citations between 1981 and 1997 to all papers cataloged by the Institute for Scientific Information [351]; (c) a 300 million vertex subset of the World Wide Web, *circa* 1999 [74]; (d) the Internet at the level of autonomous systems, April 1999 [86]; (e) the power grid of the western United States [416]; (f) the interaction network of proteins in the metabolism of the yeast *S. Cerevisiae* [212]. Of these networks, three of them, (c), (d) and (f), appear to have power-law degree distributions, as indicated by their approximately straight-line forms on the doubly logarithmic scales, and one (b) has a power-law tail but deviates markedly from power-law behavior for small degree. Network (e) has an exponential degree distribution (note the log-linear scales used in this panel) and network (a) appears to have a truncated power-law degree distribution of some type, or possibly two separate power-law regimes with different exponents.

Example: Co-purchase network



Random graphs

- Conceptual construct for modeling networks
- Simplest abstraction: a graph is drawn from a set of possible graphs according to some distribution
- Simplest less precise abstraction:
  - Each instance is described in terms of a degree distribution
    $$q(x): \text{fraction of nodes with degree } x$$
  - Need independence assumptions, a construction process
  - Power-law networks: $q(x) = x^{-\alpha}$
  - Neighbor degree distribution: $\hat{q}(x) = \frac{xq(x)}{\sum kq(k)}$
- This abstraction facilitates the development of powerful theoretical models of complex networks with “local information”

Networks and theory: Examples

Graphical Economics (Kearns and others, 2004)

- Recognize that interactions between trading partners are often local, and the nature of what is “local” can be described by an underlying network.
- Reformulate a simplified version of the Arrow-Debreu economy based on this recognition.
- Establish existence of an extension of the Arrow-Debreu equilibrium in which “local” markets clear.
- Provide a polynomial time algorithm to compute this equilibrium for a special class of underlying networks.

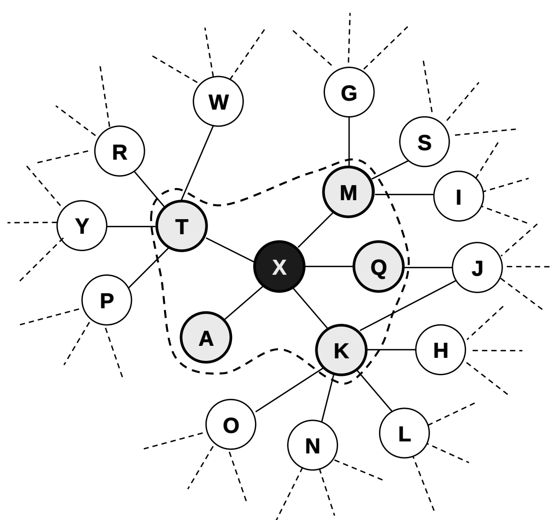
- (related: does computability of an equilibrium matter?)

Networks and theory: Examples

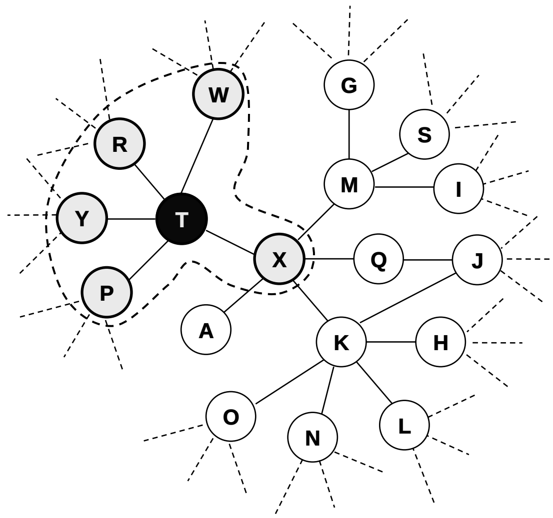
Local Network Effects (Sundararajan 2004, 2006, 2007)

- Recognizes that the value from shared interaction and adoption is often local and described by an "underlying network".
- Defines how to integrate abstractions of complex networks into an economic model whose outcome is described by a game-theoretic equilibrium.
- Establishes a homeomorphism between the standard existing solution ("fulfilled expectations" equilibria) and equilibria grounded in game theory.
- Provides the first set of properties (a Pareto-ranking, monotonicity) of the latter equilibria (a partial generalization has subsequently been provided by Galeotti et al. 2006).
- Provides a mathematical formalization of the connection between underlying networks and empirical networks.
- Shows that the optimal way to "seed" a network can often involve targeting the least connected nodes in addition to the most connected ones (and sometimes excluding the most connected ones).

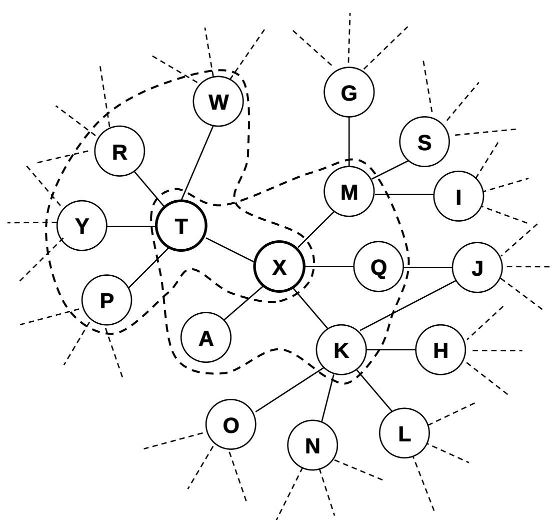
Local networks



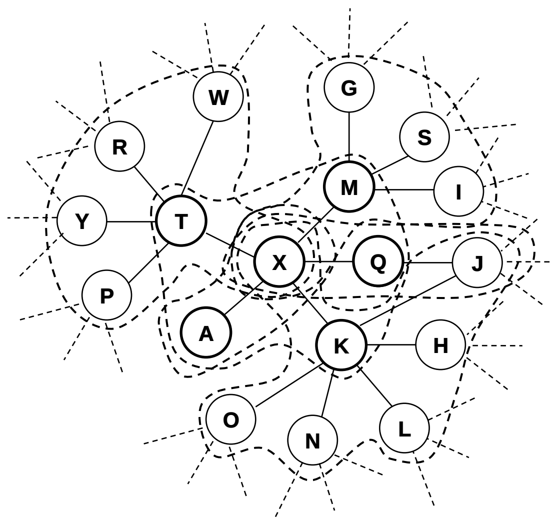
Local networks



Local networks



Local networks



Local network effects

- Agents make adoption decisions based on their observed local networks, and partial information about the entire network.
- Agents generally have:
  - different local networks
  - perfect information about the structure of their local network
  - some information about the structure of the other local networks they belong to (their neighbors' local networks)
  - very little or no information about the exact structure of the rest of the social network

A model of local network effects

- Set of potential customers $N = \{1,2,3,...,n\}$
- Single homogeneous network good that costs *c*
- Customers connected by an underlying social network modeled as an instance of a random graph (more on this soon).
- Each customer has:
  - A neighbor set G_i
  - A degree d_i (number of neighbors)
  - A valuation type θ_i (strength of adoption complementarity)
- Each customer makes an adoption choice $a_i \in \{0,1\}$
- Payoff from adoption for customer *i:*

$$a_i[u(\sum_{j \in G_i} a_j, \theta_i) - c]$$

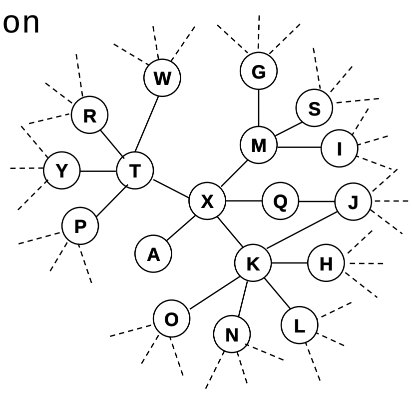
The underlying social network

$N = \{1,2,3,...,n\}$ $\quad \Gamma_i = 2^{N \setminus \{i\}}$

Set of graphs: $\Gamma \subset \Gamma_1 \times \Gamma_2 \times ... \times \Gamma_n$

Distribution over this set: $\rho : \Gamma \to [0,1]$

G: Draw from this distribution



Distribution of the social network (ρ)

For each *x* in *D*, denote

$$\Gamma_j(x) = \text{subset of } \Gamma_j \text{ such that for each } X \in \Gamma_j(x),\ \ |X| = x$$

Restrict the distribution over ρ as follows:

$$\text{For each } i, \text{for each } j \in G_i, \Pr[G_j \in \Gamma_j(x) \mid G_i, \theta_i] = q(x)$$

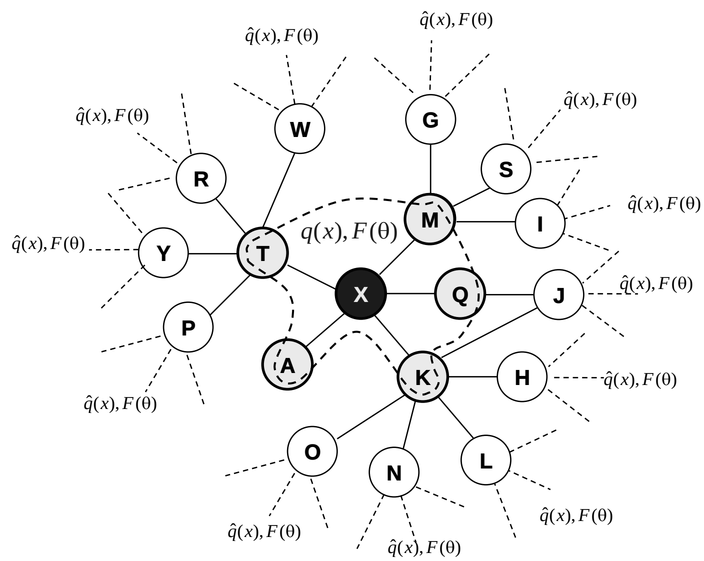
$$\text{For each } i, \text{for each } j \notin G_i, \Pr[G_j \in \Gamma_j(x) \mid G_i, \theta_i] = \hat{q}(x)$$

Generalizes to posteriors conditional on degree
Admits generalized random graphs, standard models of small world networks

Sequence of the game

- Nature draws θ_i for each *i*, draws $G \in \Gamma$
- Each agent *i* observes their type
- Each agent *i* chooses either to adopt ($a_i=1$) or not ($a_i=0$)
- Payoffs are realized

Information



Results: Equilibria

- Each symmetric Bayes-Nash equilibrium involves a threshold strategy:

$$s(d_i, \theta_i) = \begin{cases} 0, & \theta_i < \theta^*(d_i) \\ 1, & \theta_i \geq \theta^*(d_i) \end{cases}$$

with threshold $\theta^* = [\theta(1), \theta(2), ..., \theta(m)]$

- "No adoption" is always an equilibrium for pure network goods
- The equilibria can be Pareto ordered: $\Theta^* = \{\theta^A, \theta^B, ...\}$

$$\theta^A < \theta^B < ...$$

Results: Properties of the equilibria

- The ordering of equilibria is based on the equilibrium probability of neighbor adoption

$$\lambda(\theta) = \sum_{x=1}^{m} q(x)[1 - F(\theta(x))]$$

- "Higher" equilibria strictly Pareto-dominate lower ones, and therefore, there is a best equilibrium, which has the highest value of $\lambda(\theta^*)$

- Each fulfilled expectations outcome with a local expectation λ of neighbor adoption has a corresponding Bayes-Nash equilibrium with $\lambda(\theta^*) = \lambda$
  - Coordinating adoption may be simpler if it is (a) local and (b) based on a simple parameter

- Greatest equilibrium is "weakly" coalition proof: establishes a basis for stability in a standard model

The structure of adoption networks

Consider a generalized random graph with degree distribution $q(x)$, and moment generating function (MGF)

$$\Phi_p(w) = \sum_{x=0}^{\infty} q(x) w^x$$

For identical θ, and for a threshold degree δ^*, the MGF of the degree distribution of the adoption network is

$$\Phi_\alpha(w) = \Phi_p[1 - \bar{Q}(\delta^*) + w\bar{Q}(\delta^*)]$$

where

$$\bar{Q}(x) = \Pr[d_j \geq x \mid j \in G_i] = \sum_{j=x}^{m} q(x)$$

Networks and theory: Examples

Networks and Public Goods (Bramoulle & Kranton 2006)

- Theorize that knowledge gained from "costly search" is disseminated to a set of neighbors. Neighborhoods are defined by an underlying undirected network.
- The key insight: This kind of knowledge is a public good, but only locally. If an agent has a high degree, his or her effort towards searching is socially beneficial. However, agents with higher degree have a lower incentive to search because they are more connected, and are thus more likely to acquire the knowledge costlessly from a neighbor.

Why is this related to e-commerce?

- Collaborative filtering, perhaps?

Networks and theory: Examples

Networks and Social Collateral (Mobius & Szeidl 2007)

- Theorize that knowledge gained from prior commercial interaction can be transferred. The extent and reliability of transfer is mediated by an underlying network of "trust".
- The key insight: This kind of transfer is welfare improving. If an agent has a high degree, it is more likely that such transfer is viable, since the agent is more trusted. In addition, an agent who is not as connected, but whose local network is more clustered can achieve similar viable transfer, since there is better "shared" trust.

Why is this related to e-commerce?

- Reputation systems, perhaps?

Challenge: diffusion in networks

- Current dynamic models are all rooted in a baseline model of percolation on a graph.
- Probability of being "switched on" a function of how many neighbors are "on".
- SIR model
  - Equilibrium cluster distributions when an infectious disease spreads.
- SIS model
  - Approximate solutions to the cluster distribution.
- The Watts "information cascades" model.

- The output of these models tends to be a "steady state" and the time dynamics are hard to characterize.
- Lopez-Pintado et al. and Jackson/Yariv provide some integration of economic ideas, but only towards a steady-state.
- Major open question/direction for conceptual work: better models of the dynamics of diffusion of anything on a network.

Network structure and dynamic adoption

- Fixed underlying social network structure (varies between pure random and pure lattice), durable good.
- Myopic customers: adopt if their period (or myopic discounted future) value is higher than period price.
- A set of initial adopters is randomly chosen.
- Adoption proceeds until nobody adopts.

Problem (A)

- Monopoly seller of a single product, sets a price each period
- What is the optimal price path, adoption path, and how does it depend on the structure of the social network?

Network structure and dynamic adoption

- Fixed underlying social network structure (varies between pure random and pure lattice), durable good.
- Myopic customers: adopt if their period (or myopic discounted future) value is higher than period price.
- A set of initial adopters is randomly chosen.
- Adoption proceeds until nobody adopts.

Problem (B)

- Monopoly seller of a single product, sets a price each period
- Customers “pay attention” only if someone they are connected to has adopted:
  - In the prior period (the “LinkedIn” model)
  - In any prior period (the “persistent peer”, “Amway” model)
- What is the optimal price path, adoption path, and how does it depend on the structure of the social network?

Network structure and dynamic adoption

- Fixed underlying social network structure (varies between pure random and pure lattice), durable good.
- Myopic customers: adopt if their period (or myopic discounted future) value is higher than period price.
- A set of initial adopters is randomly chosen.
- Adoption proceeds until nobody adopts.

Problem (C)

- Two competing sellers of ex-ante identical goods
- Sellers choose a constant price, fraction of initial adopters
- What are the equilibrium prices and fractions?
- What is the equilibrium adoption path?
- How clustered does the network have to be to support multiple firms with similar market shares?

Network structure and dynamic adoption

- Rather than starting with no adopters, suppose a subset $S^t \subset N$ of agents are "seeded" (randomly?)
- Define

  and

$$G_i^t = G_i \cap S^t$$

$$d_i^t = | G_i |$$

- Assume that an agent knows which of its neighbors is already an adopter
- The strategy of an agent now depends on both degree as well as number of neighbors who are already adopters (that is, on both d_i^t and d_i)
- Therefore, each agent needs a posterior on both d_j^t and d_j for each $j \in G_i^t$

Network structure and dynamic adoption

- Each symmetric Bayes-Nash equilibrium involves a threshold strategy:

$$s(d_i, d_i^t, \theta_i) = \begin{cases} 0, & \theta_i < \theta^*(d_i^t, [d_i - d_i^t]) \\ 1, & \theta_i \geq \theta^*(d_i^t, [d_i - d_i^t]) \end{cases}$$

- The threshold $\theta^*(d_i^t, [d_i - d_i^t])$ is non-decreasing in both its arguments
- This result holds for any arbitrary iid posterior on the degree and adopter distribution of each $j \in G_i^t$

Network structure and dynamic adoption

- From experiments with non-strategic agents

$$s(d_i, d_i^t, \theta_i) = \begin{cases} 0, & \theta_i < \theta^*(d_i^t) \\ 1, & \theta_i \geq \theta^*(d_i^t) \end{cases}$$

  - The price path is non-monotonic over time (often tends to increase and then decrease, but not always)
  - Social networks with “small world” properties
    - take longer to get to complete adoption
    - yield higher profits

  than social networks that are more random

(2) Modeling for prediction using networked data

Modeling for prediction using networked data

Goals of this part of the tutorial

- In the short amount of time that we have, it is impossible to cover comprehensively the vast amount of related work (see bibliography for a sample)
- We will:
  - describe the four most important differences between traditional predictive modeling and predictive modeling with networked data.
  - describe example techniques and provide pointers into the literature to learn more
  - illustrate with some experiments and successful applications
- ***Considerable power for predictive inference is inherent in the structure of many networks.***

Prediction in networked data

- This part of the tutorial considers the task of modeling network data with the goal of estimating some variable
  - whose value currently is unknown
  - whose value may be categorical or numeric
  - the goal may be to estimate the value or a probability distribution over possible values
- This may be a past, current, or future value.
  - was this account defrauded?
  - is this web page of interest?
  - will this consumer respond positively to this offer?
- This will be called “prediction” to differentiate this sort of modeling from modeling with the primary goal being explanation

Prediction tasks in networked data

(cf. Getoor Tutorial 2005)

- Generic network prediction tasks
  - Node attribute value prediction
  - Node classification (special case of foregoing)
  - Link attribute value prediction
  - Predicting link existence
  - Link cardinality estimation (e.g., who's popular?)
  - Entity Resolution (e.g., is this a guy who defaulted before?)
  - Group Detection
- Related interesting network-data mining tasks
  - Graph clustering
  - Subgraph/substructure discovery
  - Finding patterns in graphs
  - see resources at end of slides

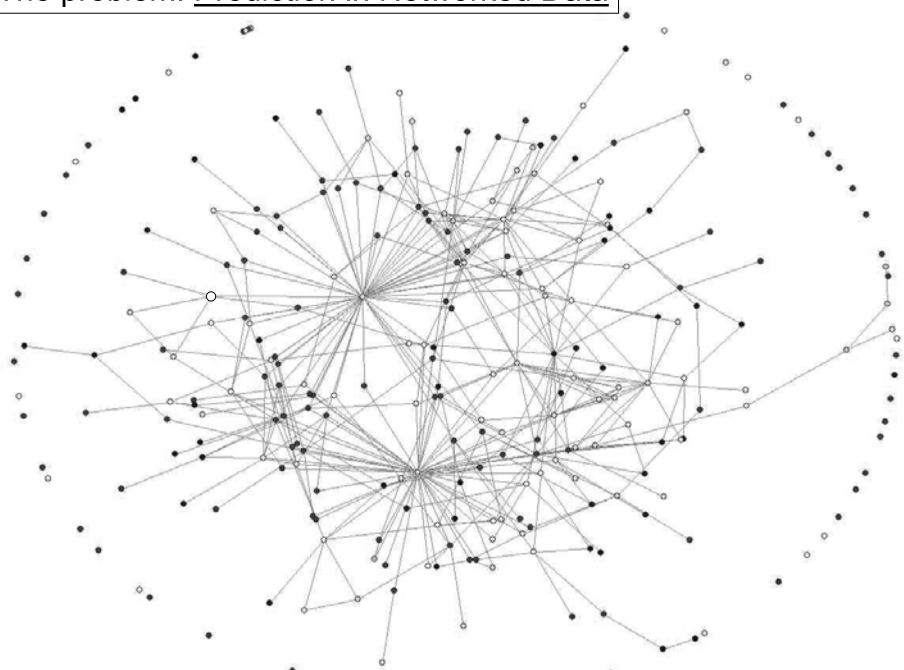
Modeling for prediction

- We assume a basic knowledge of modeling for prediction, as is done typically in applied statistics and machine learning.
- Typical techniques include:
  - linear/logistic regression, classification and regression trees, support vector machines, ensemble models (bagging, boosting, etc.), nearest-neighbor methods, neural networks, and so on.
- For background, please see:
  - Hastie, et al. (2001)
  - Mitchell (1997)

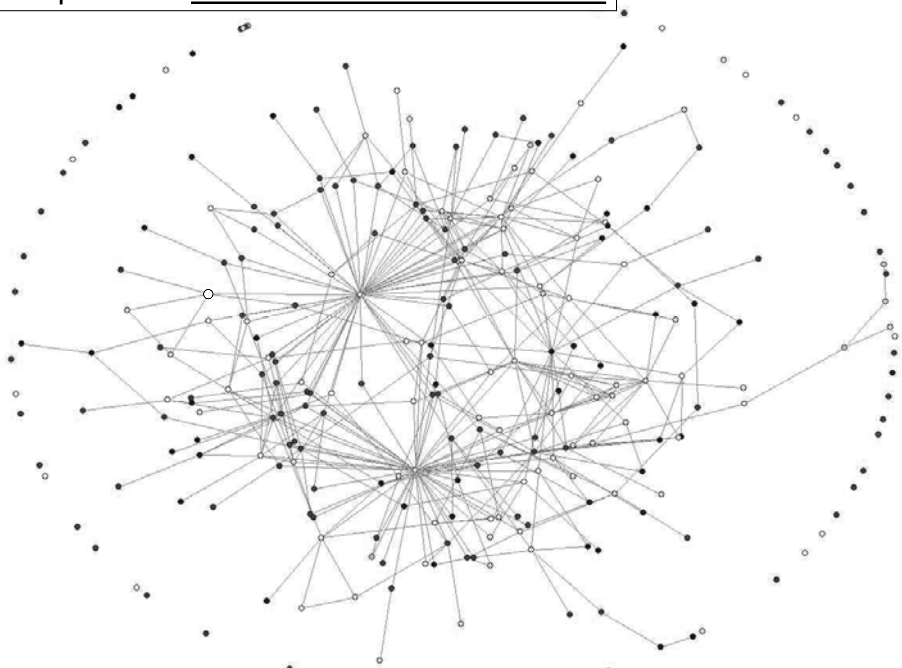
Table of Topics (perhaps incomplete)

- univariate network modeling
- network autocorrelation
  - homophily, guilt-by-association
- network feature construction
- random fields (Markov, Gaussian, Conditional)
- collective inference
  - belief propagation, MCMC, relaxation, iterative classif., graph cuts
- first-order logic modeling
- probabilistic (relational) graphical models
- combining logical and probabilistic modeling
- incorporating node identifiers
- aggregation

The problem: Prediction in Networked Data



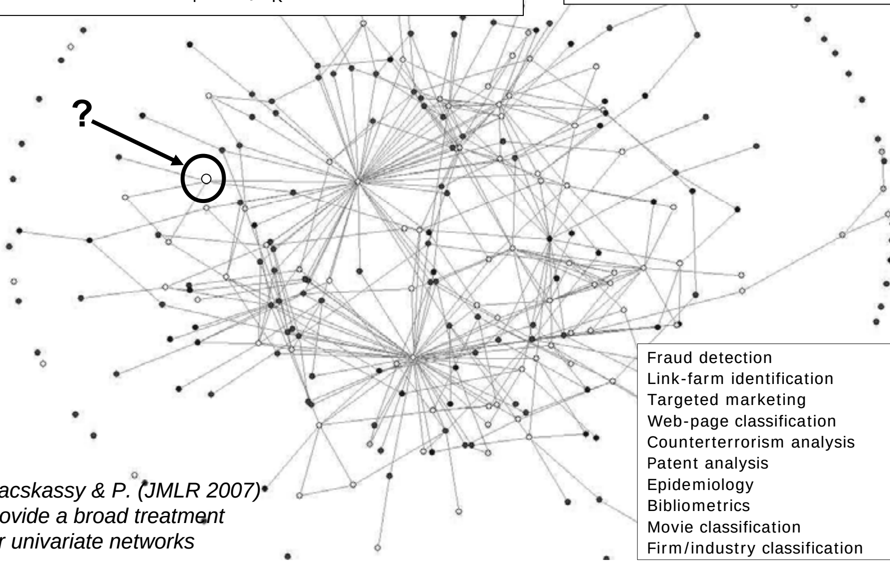
The problem: <u>Prediction in Networked Data</u>



The problem: Prediction in Networked Data

Here we'll focus on the following prediction problem:
For any node i, variable y_i, and value c, estimate $p(y_i = c|\Delta_K)$

Δ_K is everything known about the network



Macskassy & P. (JMLR 2007) provide a broad treatment for univariate networks

- Fraud detection
- Link-farm identification
- Targeted marketing
- Web-page classification
- Counterterrorism analysis
- Patent analysis
- Epidemiology
- Bibliometrics
- Movie classification
- Firm/industry classification

Example social network application:

Ecommerce firms increasingly are collecting data on explicit social networks of consumers

msn Hotmail · dialpad · Linked in Relationships Matter · Gmail talk BETA · facebook · myspace France BETA · friendster

FT.com FINANCIAL TIMES — COMPANIES Telecoms
FT Home > Companies > By sector > Telecoms

Microsoft to enter internet telephony race
By Richard Waters in San Francisco
Published: August 31 2005 02:22 | Last updated: August 31 2005 02:22

Microsoft is preparing to introduce an internet telephone service allowing calls from PCs to fixed-line or mobile telephones, extending the rapid advances by internet rivals such as Yahoo and Google into the communications business.

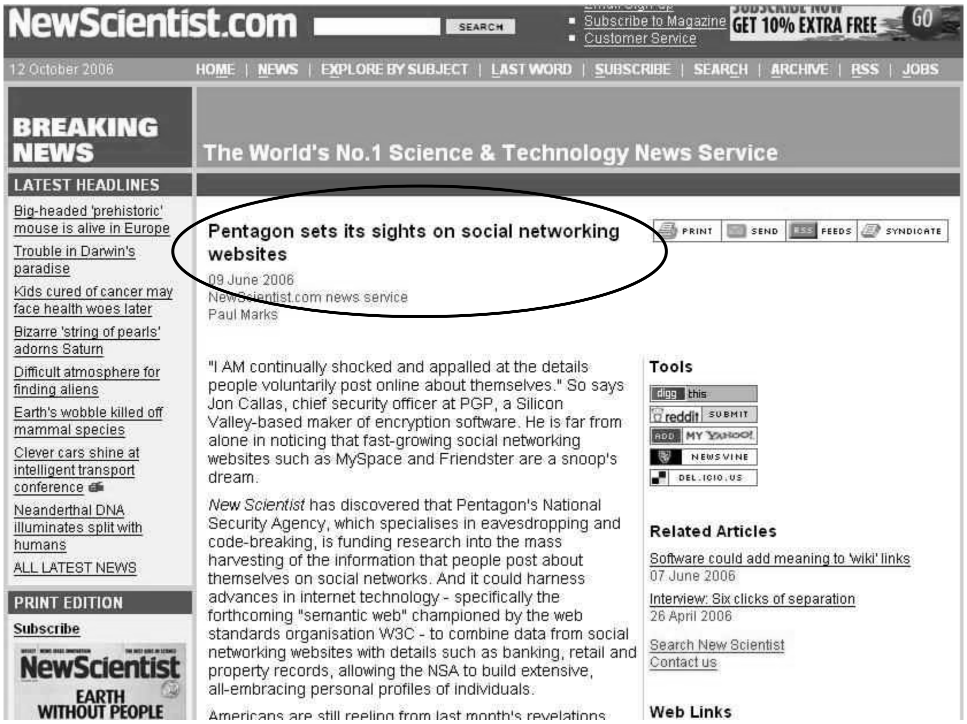
The software company will on Wednesday announce the acquisition of Teleo, a small private company whose voice-over-IP (VoIP) technology will extend the range of Microsoft's existing internet communications services. The deal echoes the acquisition by Yahoo two months ago of Dialpad and comes a week after **Google** launched a service called Google Talk that connects users over the PC.

BBC NEWS — BBC News in video and audio
News Front Page · Africa · Americas · Asia-Pacific · Europe · Middle East · South Asia · UK · Business · Market Data

Last Updated: Monday, 12 September 2005, 11:33 GMT 12:33 UK
E-mail this to a friend · Printable version

EBay to buy Skype in $2.6bn deal

Online auction site eBay has agreed to buy internet telephone company Skype


NewScientist.com SEARCH
Subscribe to Magazine
Customer Service
GET 10% EXTRA FREE GO
12 October 2006
HOME | NEWS | EXPLORE BY SUBJECT | LAST WORD | SUBSCRIBE | SEARCH | ARCHIVE | RSS | JOBS
BREAKING NEWS
The World's No.1 Science & Technology News Service
LATEST HEADLINES
Big-headed 'prehistoric' mouse is alive in Europe
Trouble in Darwin's paradise
Kids cured of cancer may face health woes later
Bizarre 'string of pearls' adorns Saturn
Difficult atmosphere for finding aliens
Earth's wobble killed off mammal species
Clever cars shine at intelligent transport conference
Neanderthal DNA illuminates split with humans
ALL LATEST NEWS
PRINT EDITION
Subscribe
NewScientist
EARTH WITHOUT PEOPLE

Pentagon sets its sights on social networking websites
09 June 2006
NewScientist.com news service
Paul Marks

PRINT SEND FEEDS SYNDICATE

"I AM continually shocked and appalled at the details people voluntarily post online about themselves." So says Jon Callas, chief security officer at PGP, a Silicon Valley-based maker of encryption software. He is far from alone in noticing that fast-growing social networking websites such as MySpace and Friendster are a snoop's dream.

New Scientist has discovered that Pentagon's National Security Agency, which specialises in eavesdropping and code-breaking, is funding research into the mass harvesting of the information that people post about themselves on social networks. And it could harness advances in internet technology - specifically the forthcoming "semantic web" championed by the web standards organisation W3C - to combine data from social networking websites with details such as banking, retail and property records, allowing the NSA to build extensive, all-embracing personal profiles of individuals.

Americans are still reeling from last month's revelations

Tools
digg this
reddit SUBMIT
ADD MY YAHOO!
NEWSVINE
DEL.ICIO.US

Related Articles
Software could add meaning to 'wiki' links
07 June 2006
Interview: Six clicks of separation
26 April 2006
Search New Scientist
Contact us

Web Links

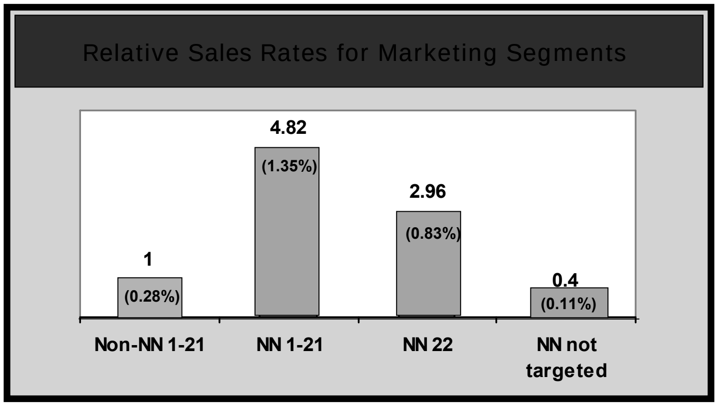
Example social network application:

Target consumers for new product

- Product: new communications service
- Long experience with targeted marketing
- Sophisticated segmentation models based on data and intuition
    - e.g., demographic, geographic, loyalty data
    - e.g., intuition regarding the types of customers known or thought to have affinity for this type of service

Hill, S., F.P., and C. Volinsky. “Network-based Marketing: Identifying likely adopters via consumer networks. ” *Statistical Science* 21 (2) 256–276, 2006.

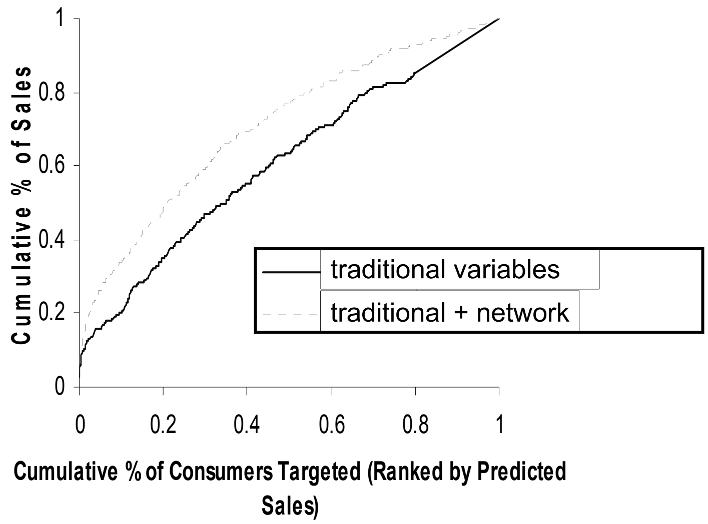
Sales rates are substantially higher for "network neighbors"



More-sophisticated network-based attributes?

Attribute	Description
Degree	Number of unique customers communicated with before the mailer
# Transactions	Number of transactions to/from customers before the mailer
Seconds of communication	Number of seconds communicated with customers before mailer
Connected to influencer?	Is an influencer in your local neighborhood?
Connected component size	Size of the connected component target belongs to.
Similarity (structural equivalence)	Max overlap in local neighborhood with existing customer

Ranking of “network neighbor” targets including more-sophisticated network-based attributes



So, what’s different about networked data?

Unique Characteristics of Networked Data (for predictive inference)

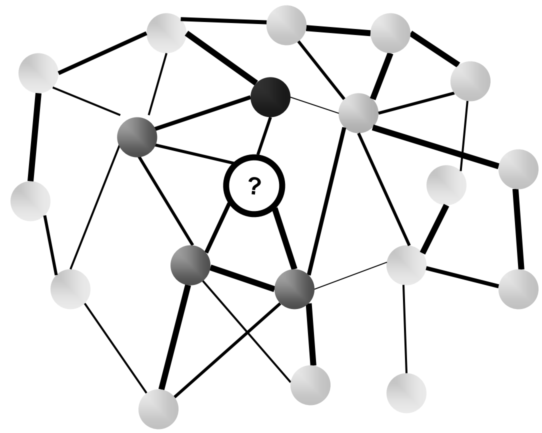
1. "Labeled" entities linked to "unlabeled" entities
   - allows "guilt-by-association" and related techniques
   - autocorrelation among neighbors
2. Collective inference is possible
   - inferences about entities can affect each other
3. Other aspects of neighbors can affect inferences about an entity
4. Identifiers can play an important role in modeling
   - being connected to specific individuals can be telling

Unique Characteristics of Networked Data (for predictive inference)

- ➢ "Labeled" entities linked to "unlabeled" entities
  - allows "guilt-by-association" and related techniques
  - autocorrelation among neighbors

2. Collective inference is possible
   - inferences about entities can affect each other
3. Other aspects of neighbors can affect inferences about an entity
4. Identifiers can play an important role in modeling
   - being connected to specific individuals can be telling

Guilt by association: autocorrelation relationship between labels* of neighboring nodes



*a label here being the value of some variable of interest

How can predictive models incorporate network autocorrelation? (Part 1)

- Features can be constructed that represent "guilt" of a node's neighbors:

$$\hat{y} = f(\ldots x_G \ldots)$$

where x_G is a (vector of) network-based guilt feature(s)

- In our network-based marketing example (Hill et al. 2006a)
  - a variable was constructed to represent whether a social-network neighbor currently uses the service.
  - And more sophisticated variables help even more.
- In fraud detection
  - variables can represent the degree to which an account is connected (via "coreference' or "cocitation" links) to known fraudulent accounts (Fawcett & P., 1997)
  - or the similarity in immediate network to known fraudulent accounts (Cortes, et al. 2001; Hill et al. 2006b)
- In hypertext classification
  - variables can be constructed representing (aggregations of) the classes of linked pages/documents (Chakrabarti et al. 1998; Lu & Getoor 2003)

Some univariate network classification techniques (see Macskassy & P. *JMLR* 2007)

- network-only Bayesian classifier nBC
  - Inspired by (Charabarti et al. 1998)
  - multinomial naïve Bayes on the neighboring class labels
- network-only link-based classifier
  - Inspired by (Lu & Getoor 2003)
  - logistic regression based on a node's "distribution" of neighboring class labels, $D_N(v_i)$ (multinomial over classes)
- relational-neighbor classifier (weighted voting)
  - (Macskassy & P. 2003, 2007)
  - More on this later

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$

- relational-neighbor classifier (class distribution)
  - Inspired by (Perlich & P. 2003)

$$p(y_i = c \mid N_i) = sim(D_N(v_i), Dist(c))$$

How can predictive models incorporate network autocorrelation? (Part 2)

- Treat network as a **random field**
  - a probability measure over a set of random variables $\{X_1, \ldots, X_n\}$ that gives non-zero probability to any configuration of values for all the variables.
- Convenient for modeling network data

  A **Markov random field** satisfies:

$$p(X_i = x_i \mid X_j = x_j, i \neq j) = p(X_i = x_i \mid N_i)$$

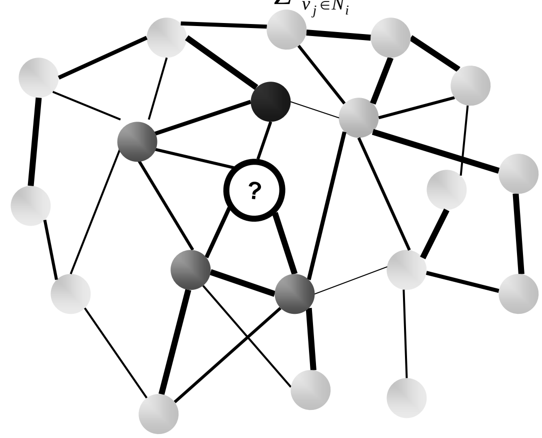
  - where N_i is the set of neighbors of X_i under some definition of neighbor.
  - in other words, the probability of a variable taking on a value depends only on its neighbors

(Dobrushin, 1968; Besag, 1974; Geman and Geman, 1984)

How can predictive models incorporate network autocorrelation? (Part 2, cont.)

A particularly simple guilt-by-association model is that a value's probability is the average of its probabilities at the neighboring nodes

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



- *Gaussian random field (Besag 1975; Zhu et al. 2003)*
- *"Relational neighbor" classifier - wvRN (Macskassy & P. 2003)*

How can predictive models incorporate network autocorrelation? (Part 2, cont.)

- Random fields have a long history for modeling regular grid data
  - in statistical physics, spatial statistics, image analysis
  - see Besag (1974)
- Besag (1975) applied such methods to what we would call networked data ("non-lattice data")
- Some notable example applications to electronic commerce applications:
  - hypertext classification (Chakrabarti et al. 1998)
  - viral marketing (Domingos & Richardson 2001)
  - eBay auction fraud (Pandit et al. 2007)

Is guilt-by-association justified theoretically?

Thanks to (McPherson, et al., 2001)

- *Birds of a feather, flock together*

  – attributed to Robert Burton (1577-1640)

- (*People) love those who are like themselves*

  -- Aristotle, *Rhetoric* and *Nichomachean Ethics*

- *Similarity begets friendship*

  -- Plato, *Phaedrus*

- *Hanging out with a bad crowd will get you into trouble*

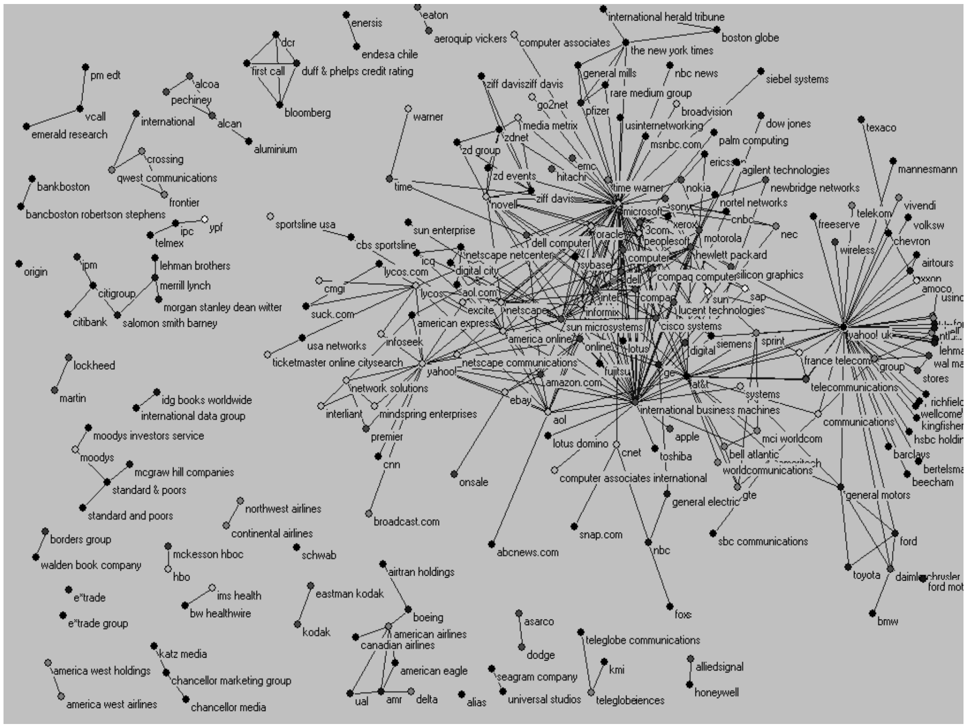
  -- Foster's Mom

Is guilt-by-association justified theoretically?

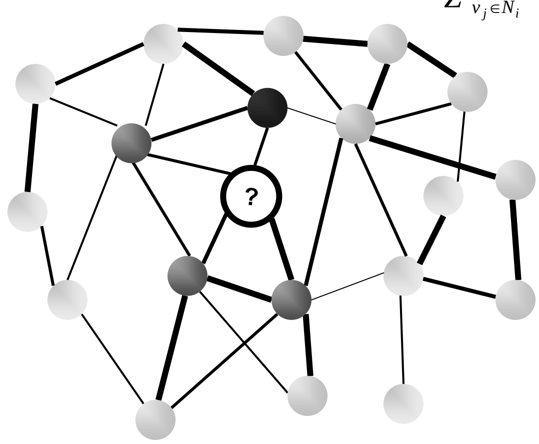
Homophily

- fundamental concept underlying social theories
  - (e.g., Blau 1977)
- one of the first features noticed by analysts of social network structure
  - antecedents to SNA research from 1920's (Freeman 1996)
- fundamental basis for links of many types in social networks (McPherson, et al., Annu. Rev. Soc. 2001)
  - *Patterns of homophily:*
  - *remarkably robust across widely varying types of relations*
  - *tend to get stronger as more relationships exist*
- Now being considered in mathematical analysis of networks ("assortativity", e.g., Newman (2003))

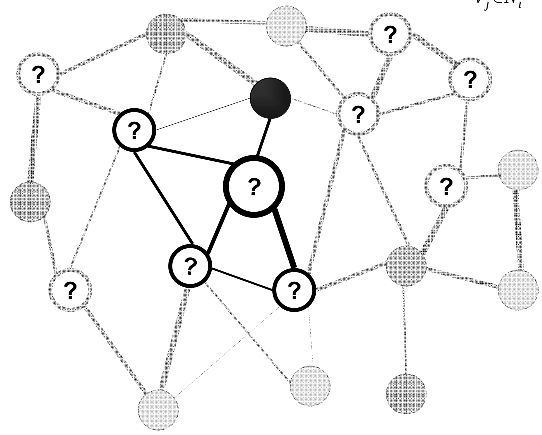
- Does it apply to non-social networks?



$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



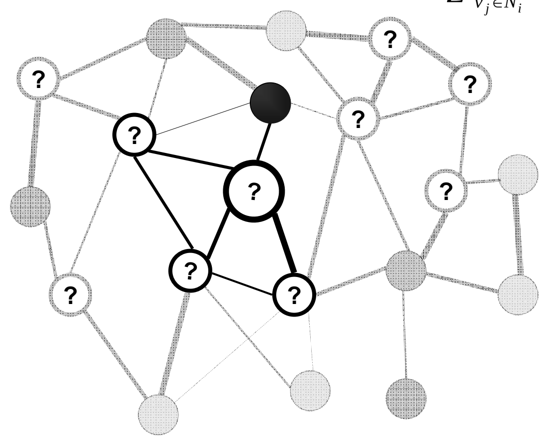
Unique Characteristics of Networked Data (for predictive inference)

1. "<u>Labeled</u>" entities linked to "unlabeled" entities
   - allows "guilt-by-association" and related techniques
   - autocorrelation among neighbors

➢ <u>Collective</u> inference is possible
   - <u>inferences</u> about entities can affect each other

3. <u>Other aspects</u> of neighbors can affect inferences about an entity
4. <u>Identifiers</u> can play an important role in modeling
   - being connected to specific <u>individuals</u> can be telling

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



Various techniques for collective inference

(see also Jensen et al. KDD 2004)

- Gibbs sampling (Geman & Geman 1984)
- Iterative classification (Besag 1986; …)
- Relaxation labeling (Rosenfeld et al. 1976; …)
- Loopy belief propagation (Pearl 1988)
- Graph-cut methods (Greig et al. 1989; …)

Either:

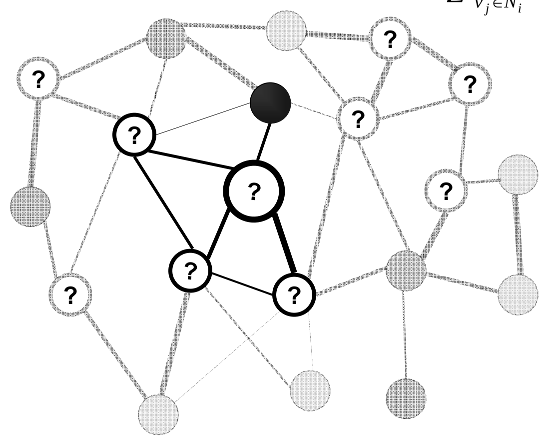
1. *estimate the maximum a posteriori joint probability distribution of all free parameters*

or

2. *estimate the marginal distributions of some or all free parameters simultaneously* *(or some related likelihood-based scoring)*

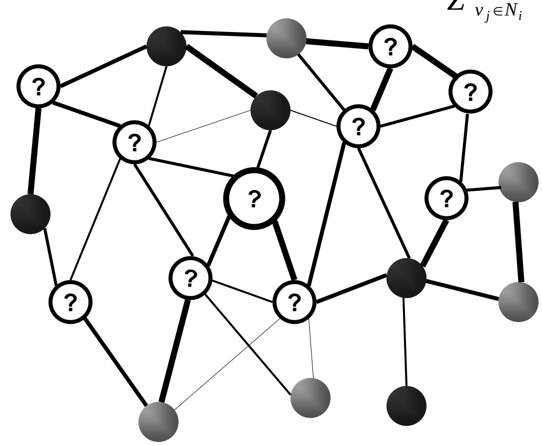
Collective inference example: iterative classification

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



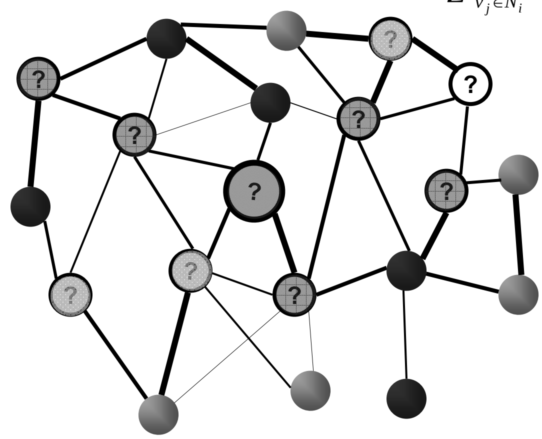
Collective inference example: iterative classification

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



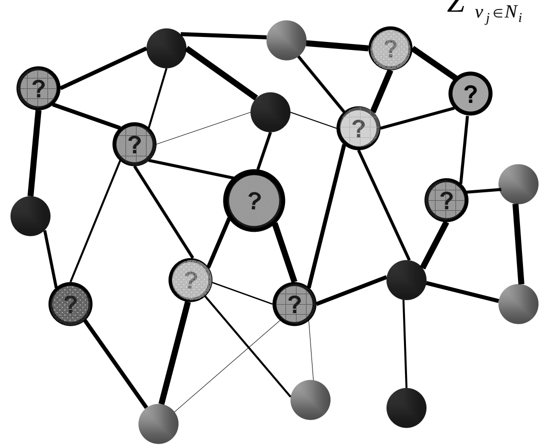
Collective inference example: iterative classification

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



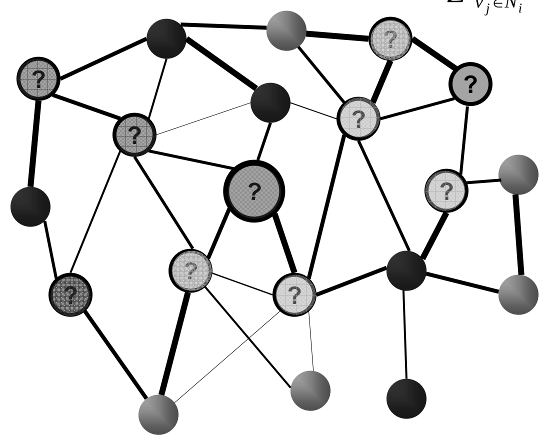
Collective inference example: iterative classification

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



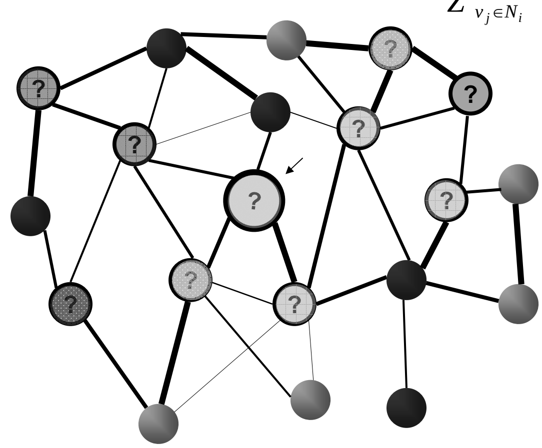
Collective inference example: iterative classification

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



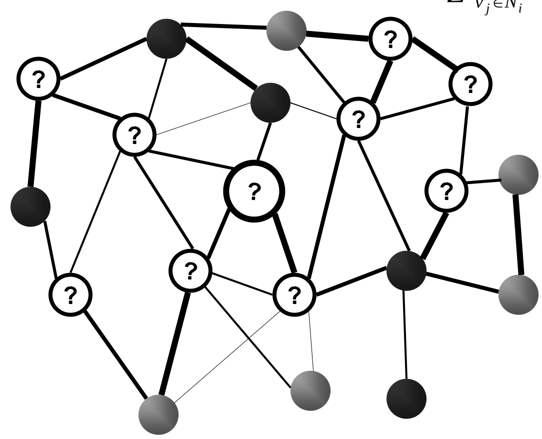
Collective inference example: iterative classification

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



Collective inference example: iterative classification

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$



- recall network-based marketing example?

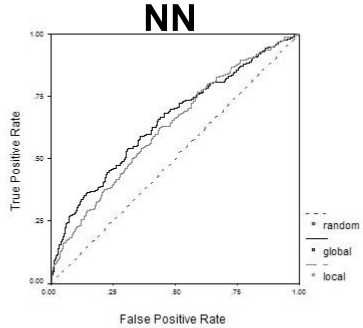
➔ collective inference can help for the nodes that are not neighbors of existing customers

Collective inference gives additional improvement, especially for non-network neighbors

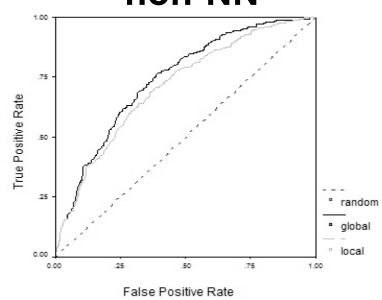
Hill et al. 2007

Attribute	NN	non-NN
All first-order network variables	0.61	0.71
All first-order + oracle (wvRN)	0.63	0.74
All first-order + CI score (wvRN)	**0.62**	**0.74**

NN



non-NN



So, how much "information" is in the network structure alone?

Network Classification Case Study

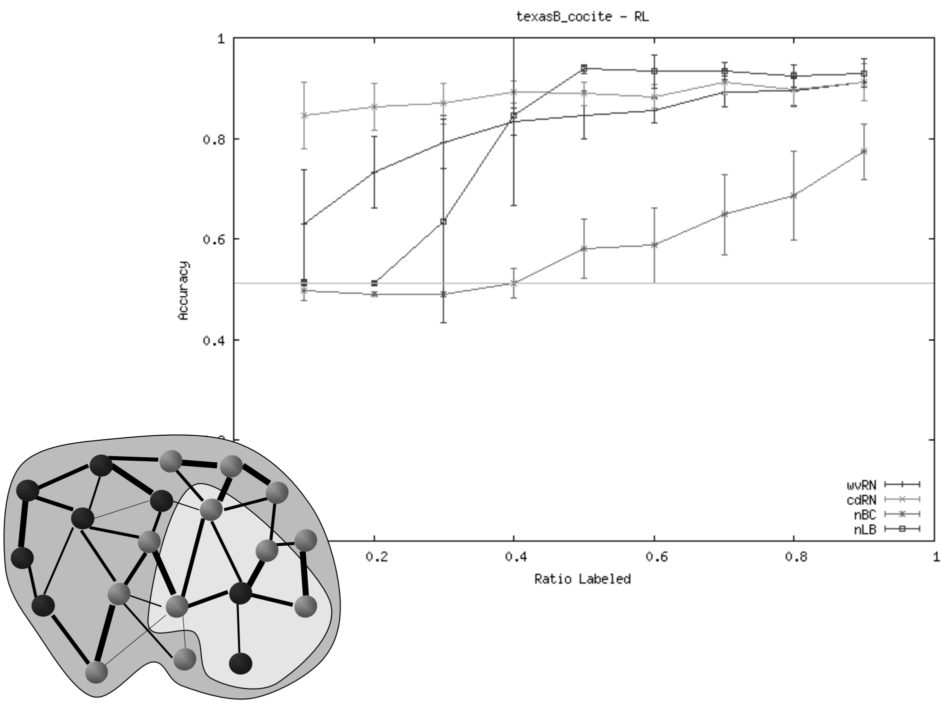
- 12 data sets from 4 domains
    - (previously used in ML research)
  - IMDB (Internet Movie Database) (e.g., Jensen & Neville, 2002)
  - Cora (e.g., Taskar et al., 2001) [McCallum et al., 2000]
  - WebKB [Craven et al., 1998]
    - CS Depts of Texas, Wisconsin, Washington, Cornell
    - multiclass & binary (student page)
    - "cocitation" links
  - Industry Classification [Bernstein et al., 2003]
    - yahoo data, prnewswire data
- Homogeneous nodes & links
  - one type, different classes/subtypes
- Univariate classification
  - only information: structure of network and (some) class labels
  - guilt-by-association (wvRN) with collective inference
  - plus several models
  - that "learn" relational patterns

Macskassy, S. and F. P. "Classification in Networked Data: A toolkit and a univariate case study." *Journal of Machine Learning Research 2007*.

How much information is in the network structure?

Data set	Accuracy	Relative error reduction over default prediction
wisconsin-student	0.94	86%
texas-student	0.93	86%
Cora	0.87	81%
wisconsin-multi	0.82	67%
cornell-student	0.85	65%
imdb	0.83	65%
wash-student	0.85	58%
wash-multi	0.71	52%
texas-multi	0.74	50%
industry-yahoo	0.64	49%
cornell-multi	0.68	45%
industry-pr	0.54	36%

- Labeling 90% of nodes
- Classifying remaining 10%
- Averaging over 10 runs



Univariate network classification techniques

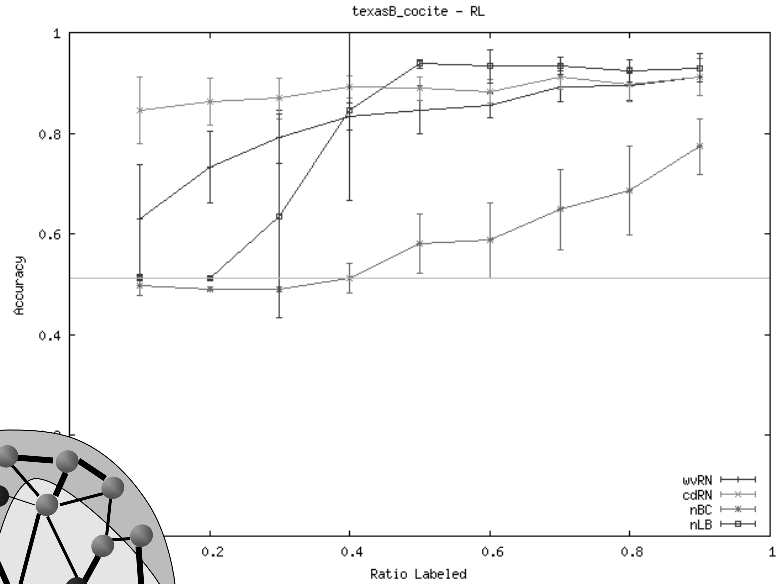
(see Macskassy & Provost 2007)

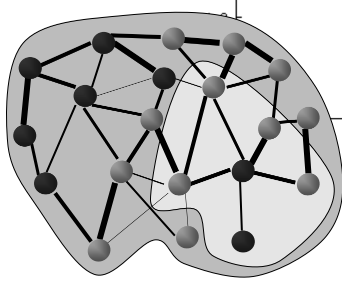
- nBC - network-only Bayesian classifier
  - Inspired by (Charabarti et al. 1998)
  - multinomial naïve Bayes on the neighboring class labels
- nLC - network-only link-based classifier
  - Inspired by (Lu & Getoor 2003)
  - logistic regression based on a node's "distribution" of neighboring class labels, $D_N(v_i)$ (multinomial over classes)
- wvRN - relational-neighbor classifier (weighted voting)
  - (Macskassy & P. 2003, 2007)

$$p(y_i = c \mid N_i) = \frac{1}{Z} \sum_{v_j \in N_i} w_{i,j} \cdot p(y_j = c \mid N_j)$$

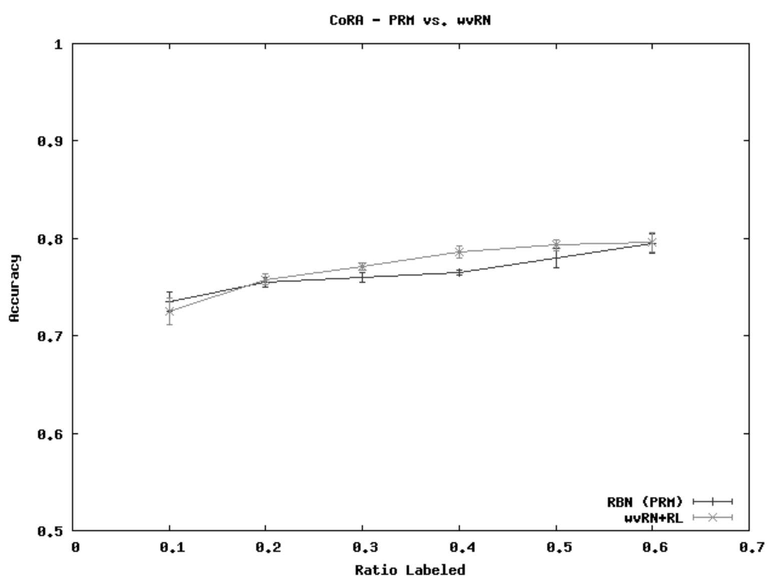
- cdRN relational-neighbor classifier (class distribution)
  - Inspired by (Perlich & P. 2003)

$$p(y_i = c \mid N_i) = sim(D_N(v_i), Dist(c))$$

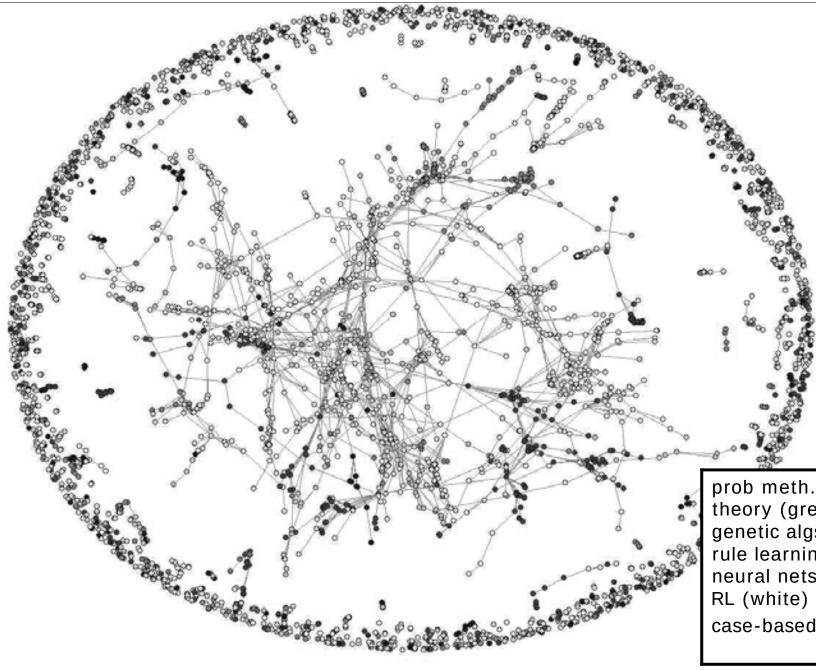




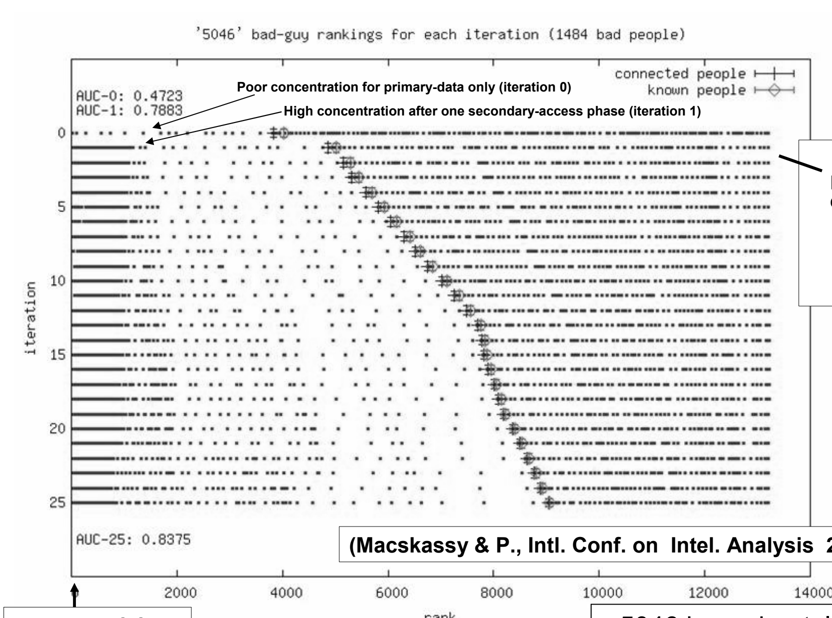
RBN vs wvRN
Classifying linked documents (CoRA)



Machine Learning Research Papers (from CoRA data)



prob meth. (yellow)
theory (green)
genetic algs (red)
rule learning (blue)
neural nets (pink)
RL (white)
case-based (orange)



rightmost people are completely unknown, therefore ranking is uniform

(Macskassy & P., Intl. Conf. on Intel. Analysis 2005)

most suspicious

- high concentration of bad guys at "top" of suspicion ranking
- gets better with increased secondary-data access

5046 is moderately noisy: ¼ of "known" bad guys were mislabeled

Unique Characteristics of Networked Data (for predictive inference)

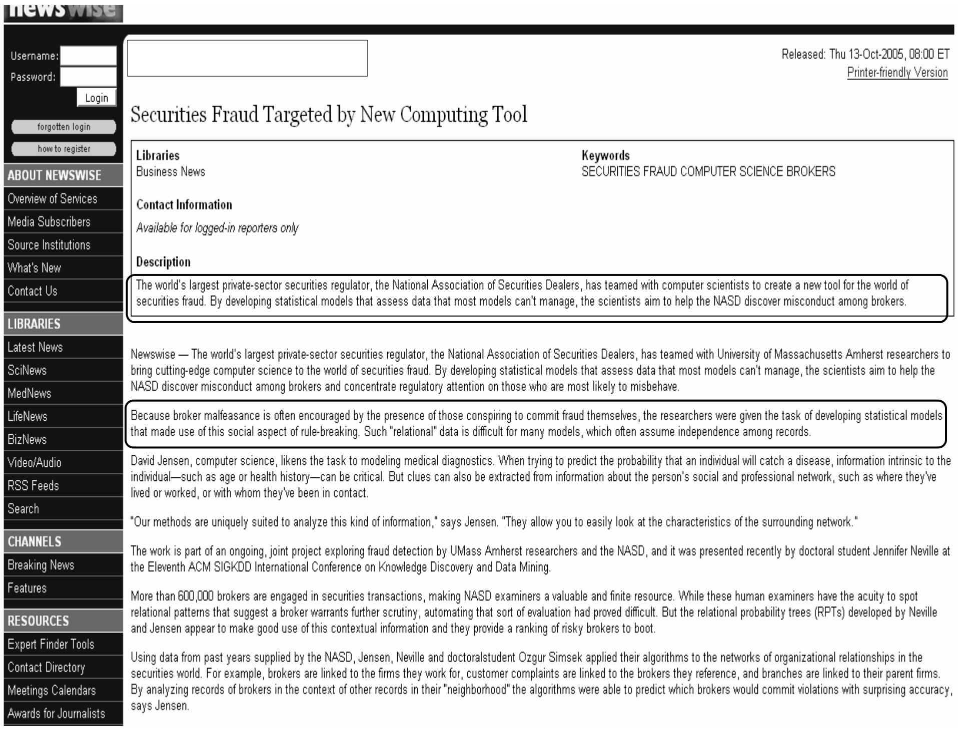
1. "Labeled" entities linked to "unlabeled" entities
   – allows "guilt-by-association" and related techniques
   – autocorrelation among neighbors
2. Collective inference is possible
   – inferences about entities can affect each other

➢ Other aspects of neighbors can affect inferences about an entity

4. Identifiers can play an important role in modeling
   – being connected to specific individuals can be telling

Networks ≠ Graphs?

- Networked data can be much more complex than just sets of (labeled) vertices and edges.
  – Vertices and edges can be heterogeneous
  – Vertices and edges can have various information associated with them
- Example: Consider the following problem
  – Can we estimate the likelihood that a stock broker is/will be engaged in activity that violates securities regulations?



newswise

Username:
Password:
Login
forgotten login
how to register

ABOUT NEWSWISE
Overview of Services
Media Subscribers
Source Institutions
What's New
Contact Us

LIBRARIES
Latest News
SciNews
MedNews
LifeNews
BizNews
Video/Audio
RSS Feeds
Search

CHANNELS
Breaking News
Features

RESOURCES
Expert Finder Tools
Contact Directory
Meetings Calendars
Awards for Journalists

Released: Thu 13-Oct-2005, 08:00 ET
Printer-friendly Version

Securities Fraud Targeted by New Computing Tool

Libraries
Business News

Keywords
SECURITIES FRAUD COMPUTER SCIENCE BROKERS

Contact Information
Available for logged-in reporters only

Description
The world's largest private-sector securities regulator, the National Association of Securities Dealers, has teamed with computer scientists to create a new tool for the world of securities fraud. By developing statistical models that assess data that most models can't manage, the scientists aim to help the NASD discover misconduct among brokers.

Newswise — The world's largest private-sector securities regulator, the National Association of Securities Dealers, has teamed with University of Massachusetts Amherst researchers to bring cutting-edge computer science to the world of securities fraud. By developing statistical models that assess data that most models can't manage, the scientists aim to help the NASD discover misconduct among brokers and concentrate regulatory attention on those who are most likely to misbehave.

Because broker malfeasance is often encouraged by the presence of those conspiring to commit fraud themselves, the researchers were given the task of developing statistical models that made use of this social aspect of rule-breaking. Such "relational" data is difficult for many models, which often assume independence among records.

David Jensen, computer science, likens the task to modeling medical diagnostics. When trying to predict the probability that an individual will catch a disease, information intrinsic to the individual—such as age or health history—can be critical. But clues can also be extracted from information about the person's social and professional network, such as where they've lived or worked, or with whom they've been in contact.

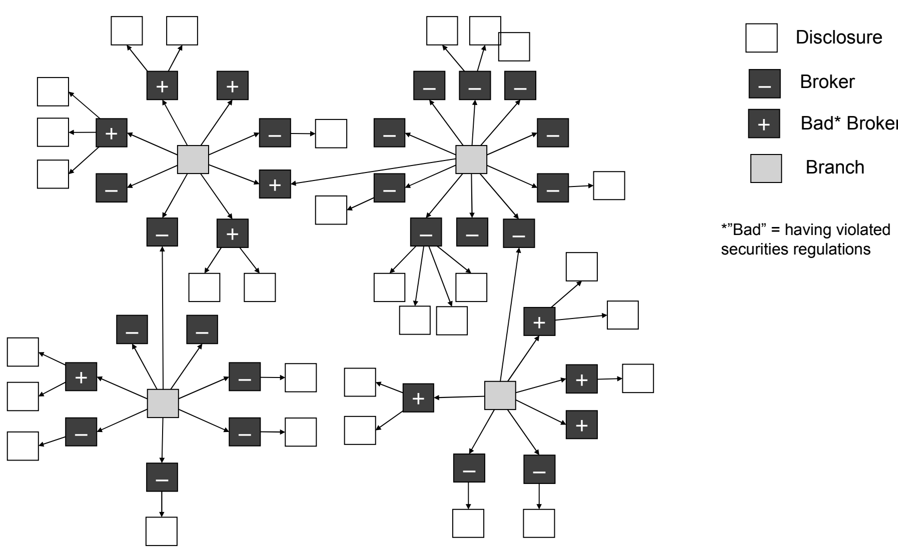
"Our methods are uniquely suited to analyze this kind of information," says Jensen. "They allow you to easily look at the characteristics of the surrounding network."

The work is part of an ongoing, joint project exploring fraud detection by UMass Amherst researchers and the NASD, and it was presented recently by doctoral student Jennifer Neville at the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

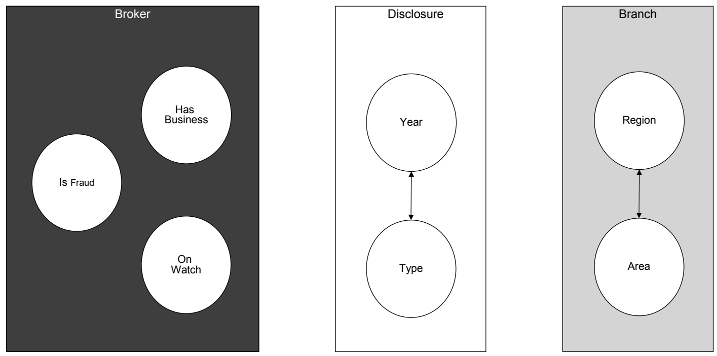
More than 600,000 brokers are engaged in securities transactions, making NASD examiners a valuable and finite resource. While these human examiners have the acuity to spot relational patterns that suggest a broker warrants further scrutiny, automating that sort of evaluation had proved difficult. But the relational probability trees (RPTs) developed by Neville and Jensen appear to make good use of this contextual information and they provide a ranking of risky brokers to boot.

Using data from past years supplied by the NASD, Jensen, Neville and doctoralstudent Ozgur Simsek applied their algorithms to the networks of organizational relationships in the securities world. For example, brokers are linked to the firms they work for, customer complaints are linked to the brokers they reference, and branches are linked to their parent firms. By analyzing records of brokers in the context of other records in their "neighborhood" the algorithms were able to predict which brokers would commit violations with surprising accuracy, says Jensen.

Detecting "bad brokers" (NASD)
(Neville et al. KDD 2005)



Data on brokers, branches, disclosures
(Neville et al. KDD 2005)



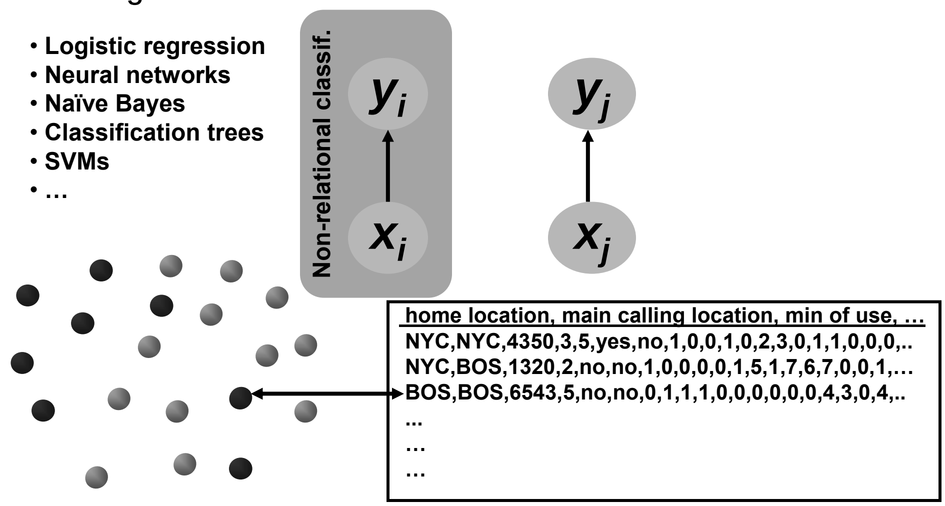
Relational Learning

- ***Relational learning and inference***: learning and inference where one cannot represent data as a single relation/table of independently distributed entities, without losing important information
- For example, data may be represented as a non-trivial, multi-table relational database, or as a heterogeneous, attributed graph, or in first-order logic.
- There is a huge literature on relational learning (see resources slide toward end for pointers) and it would be impossible to do justice to it in the short amount of time we have.
- Let's consider briefly three approaches
  - model in first-order logic
  - model as probabilistic graphical model
  - do both

Traditional Learning and Classification

Setting:

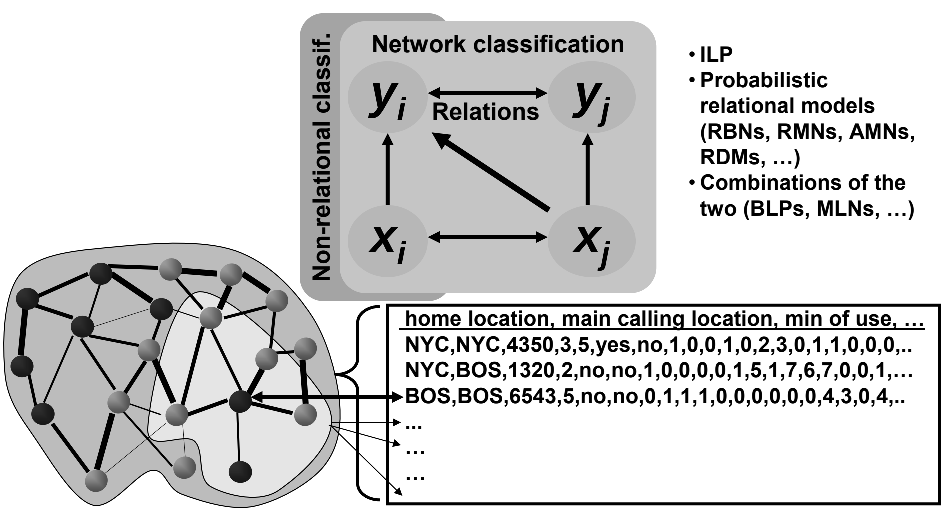
- Logistic regression
- Neural networks
- Naïve Bayes
- Classification trees
- SVMs
- …



Network Learning and Classification

Setting:

- ILP
- Probabilistic relational models (RBNs, RMNs, AMNs, RDMs, …)
- Combinations of the two (BLPs, MLNs, …)



First-order logic modeling

- The field of Inductive Logic Programming has extensively studied modeling in first-order logic, which can represent complicated relational and graph data
- Although it has been changing, traditionally ILP did not focus on representing uncertainty

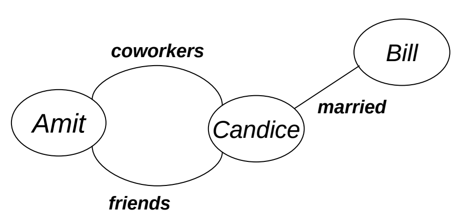
— in the usual use of first-order logic, *each ground atom either is true or is not true* (cf., a Herbrand interpretation)

…one of the reasons for the modern rubric "statistical relational learning"

- First-order logic for statistical modeling of network data?
  - a strength is its ability to represent and search for complex and deep patterns in the network
  - a weakness is its relative lack of support for aggregations across nodes (beyond *existence*)
  - more on this in a minute…

Network data in first-order logic

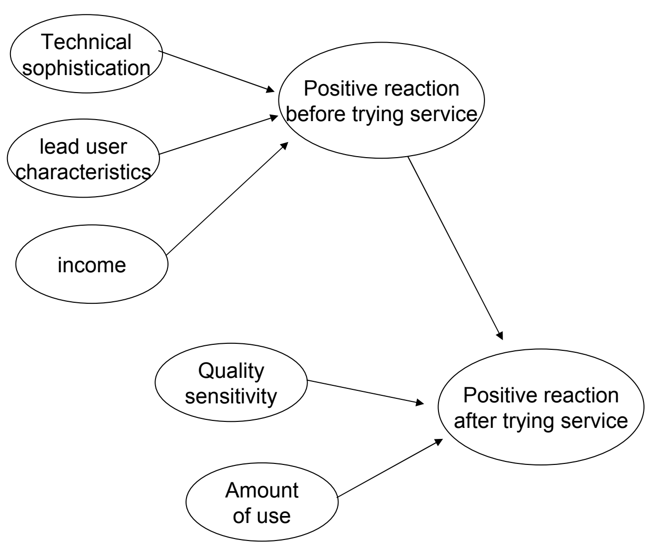
- broker(Amit), broker(Bill), broker(Candice), …
- works_for(Amit, Bigbank), works_for(Bill, E_broker), works_for(Candice, Bigbank), …
- married(Candice, Bill)
- smokes(Amit), smokes(Candice), …
- works_for(X,F) & works_for(Y,F) -> coworkers(X,Y)
- smokes(X) & smokes(Y) & coworkers(X,Y) -> friends(X,Y)
- …



Probabilistic graphical models

- Probabilistic graphical models (PGMs) are convenient methods for representation of (and inference with) probability distributions across a set of variables.
  - Bayesian networks (BNs), Markov networks (MNs), Dependency networks (DNs)
  - See Pearl (1988), Heckerman et al. (2000)
- Typically BNs, MNs, DNs are used to represent a set of random variables describing <u>independent</u> instances.
  - For example, the probabilistic dependencies among the descriptive features of a consumer—the same for different consumers

Example: A Bayesian network modeling consumer reaction to new service



Probabilistic <u>relational</u> models

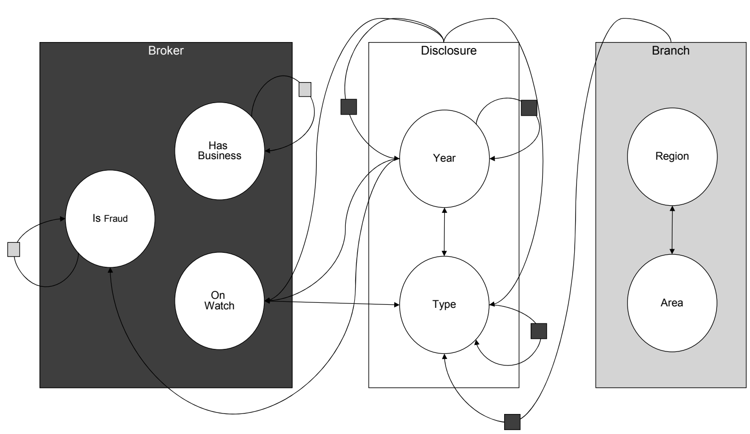
The term “relational” recently has been used to distinguish the use of PGMs to represent variables across a set of dependent, multivariate instances.

- For example, the dependencies between the descriptive features of friends in a social network
- We saw a “relational” Markov network earlier when we discussed Markov random fields for univariate network data
  - although the usage is not consistent, “Markov random field” often is used for a MN over multiple instances of the “same” variable
- RBNs (Koller and Pfeffer,1998; Friedman et al., 1999; Taskar et al., 2001), RMNs (Taskar et al. 2002), RDNs (Neville & Jensen, 200
- In th

Conditional random fields (CRFs, Lafferty et al., 2001) are random fields where the probability of a node’s label is conditioned not only on the labels of neighbors (as in MRFs), but also on all the observed attribute data.

Relational prob. model of broker variables

(Neville & Jensen, JMLR to appear)

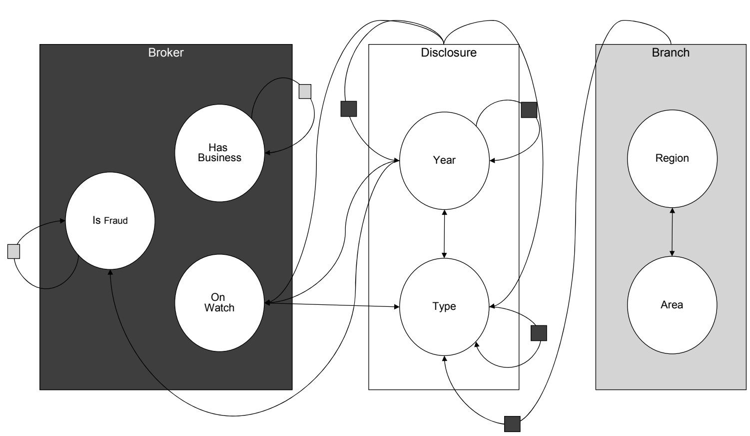


note: needs to be “unrolled” across network

Important concept!

- The network of statistical dependencies does not necessarily correspond to the data network
- Example on next three slides…

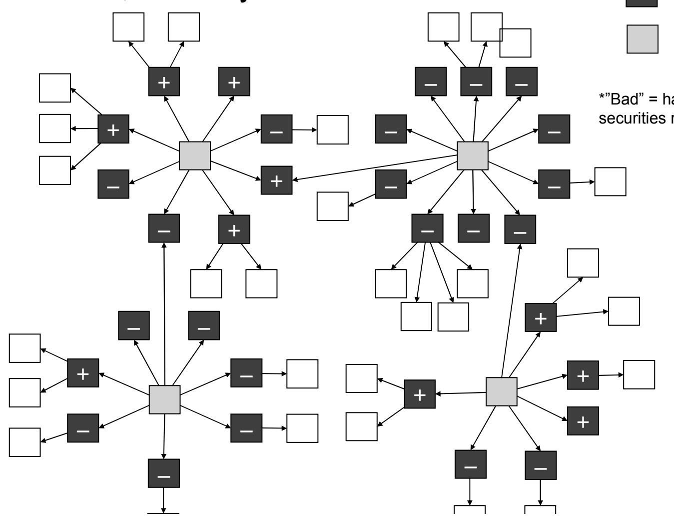
Recall: broker dependency network



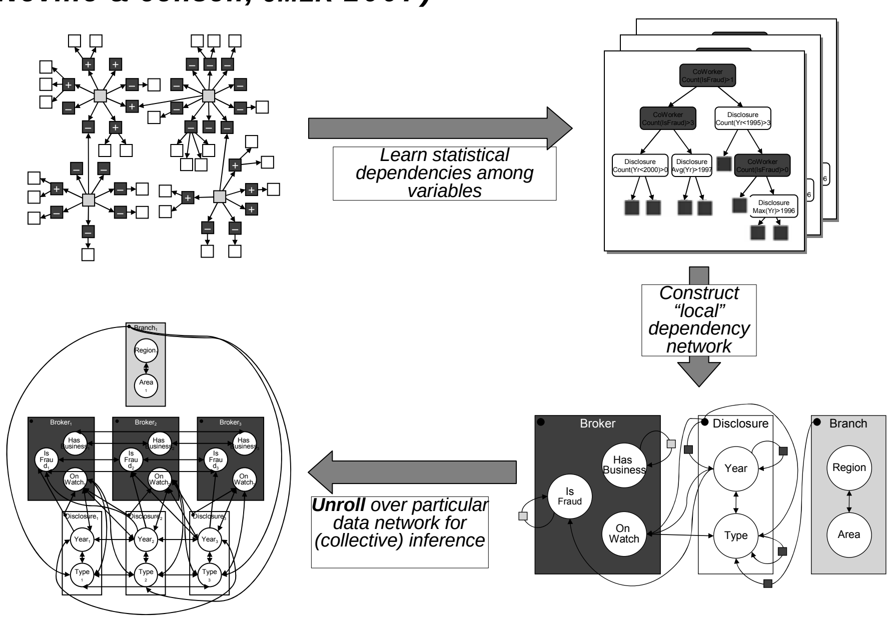
note: this dependency network needs to be "unrolled" across the data network

Broker data network
(Neville et al. 2005)

Statistical dependencies between brokers “jump across” branches; similarly for disclosures

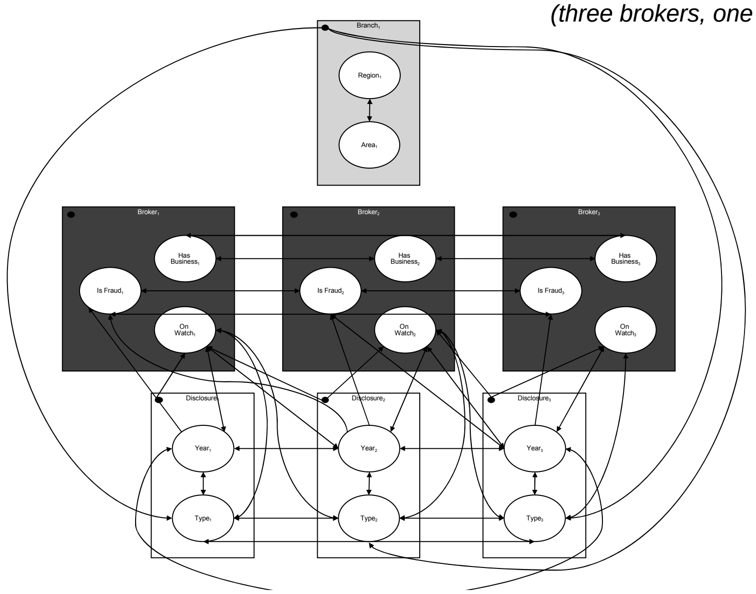


Putting it all together:
Relational dependency networks
(Neville & Jensen, JMLR 2007)



Model unrolled on (tiny) data network

(three brokers, one branch)



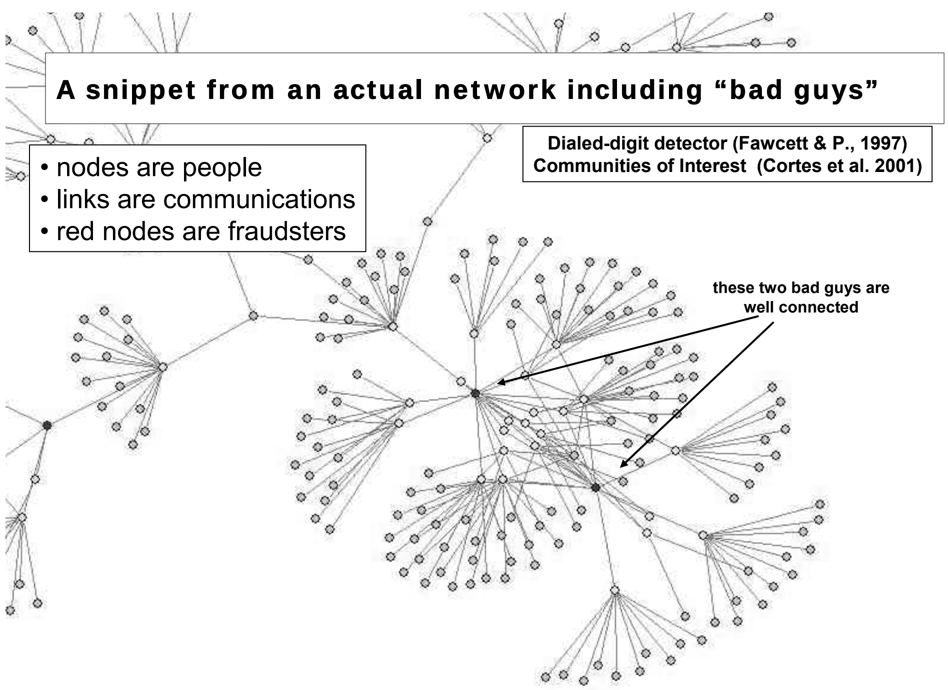
Combining first-order logic and probabilistic graphical models

- Recently there have been efforts to combine FOL and probabilistic graphical models
  - e.g., Bayesian logic programs (Kersting and de Raedt, 2001), Markov logic networks (Richardson & Domingos, MLJ 2006)
  - *and see discussion & citations in (Richardson & Domingos, 2006)*
- For example: Markov logic networks
  - A template for constructing Markov networks
    - and therefore, a model of the joint distribution over a set of variables
  - A first-order knowledge base with a weight for each formula
- Advantages:
  - Markov network gives sound probabilistic foundation
  - first-order logic allows compact representation of large networks and a wide variety of domain knowledge

Unique Characteristics of Networked Data (for predictive inference)

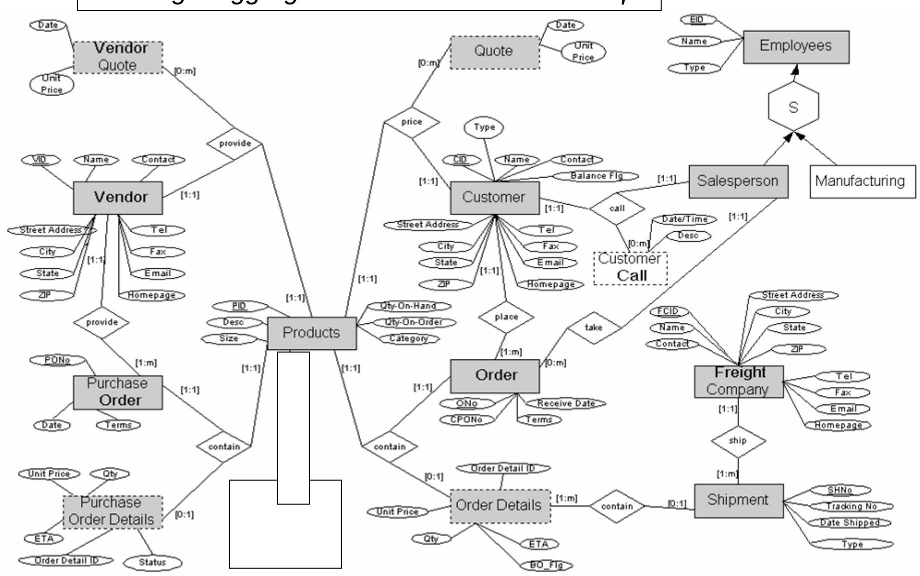
1. "Labeled" entities linked to "unlabeled" entities
   – allows "guilt-by-association" and related techniques
   – autocorrelation among neighbors
2. Collective inference is possible
   – inferences about entities can affect each other
3. Other aspects of neighbors can affect inferences about an entity

➢ Identifiers can play an important role in modeling
   – being connected to specific individuals can be telling



<u>Side note</u>: not just for "networked data" – id's important for any data in a multi-table RDB

challenge: aggregation over 1-to-n relationships



How to incorporate identifiers of related objects (in a nutshell)

1. Estimate from known data:
   - *class-conditional distributions* of related identifiers (say D^+ & D^-)
   - can be done, for example, assuming class-conditional independence in analogy to Naïve Bayes
   - save these as "meta-data" for use with particular cases
2. Any particular <u>case</u> C has its own "distribution" of related identifiers (say D_c)
3. Create features
   - $\Delta(D_c, D^+)$, $\Delta(D_c, D^-)$, $(\Delta(D_c, D^+) - \Delta(D_c, D^-))$
   - where Δ is a distance metric between distributions
4. Add these features to target-node description(s) for learning/estimation

<u>Main idea</u>:

"Is the distribution of nodes to which this case is linked similar to that of a <whatever>?"

(Perlich & P., 2006)

Density Estimation for Aggregation

1: Class-conditional distributions

Distr.	A	B
$D_{Class\ 1}$	0.75	0.25
$D_{Class\ 0}$	0.2	0.8

CID	*Class*
C1	0
C2	1
C3	1
C4	0

CID	*id*
C1	B
C2	A
C2	A
C2	B
C3	A
C4	B
C4	B
C4	B
C4	A

?

2: Case distributions:

D_c	A	B
C1	0	1
C2	0.66	0.33
C3	1	0
C4	0.25	0.75

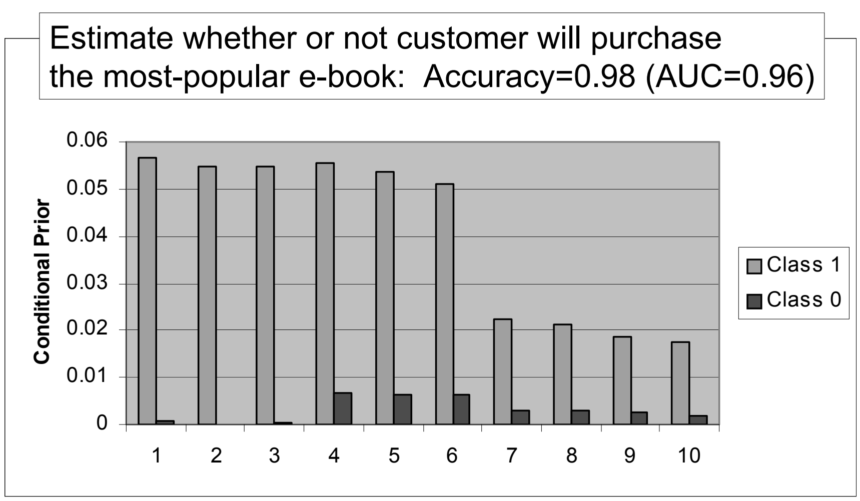
3: L2 distances for C1:

L2(C1, $D_{Class\ 1}$) = 1.125
L2(C1, $D_{Class\ 0}$) = 0.08

4: Extended feature vector:

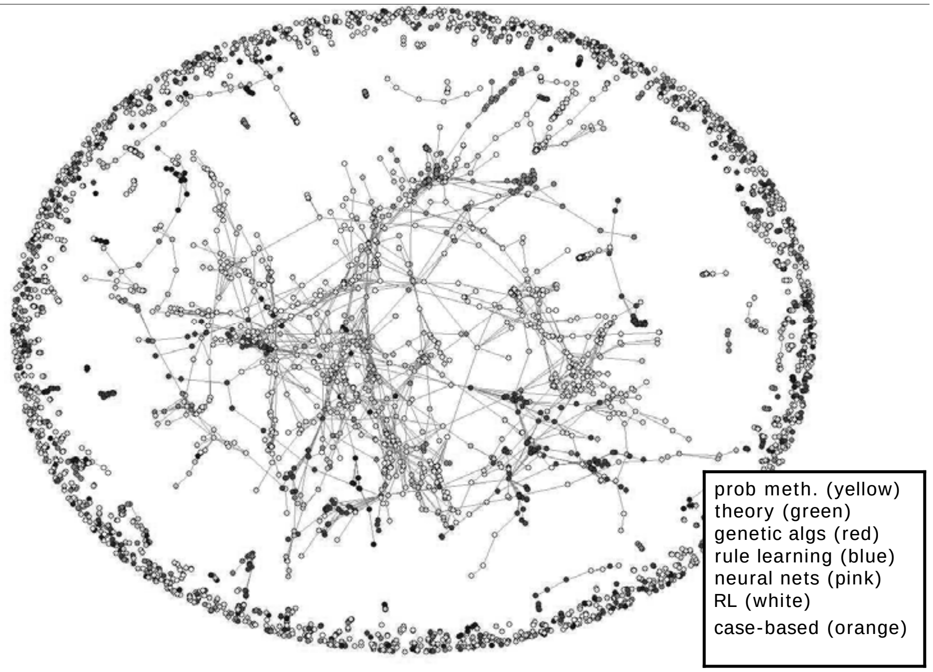
CID	*$L2_1$*	*$L2_0$*	*$L2_1 - L2_0$*	*Class*
C1	1.125	0.08	-1.045	0
C2	0.014	0.435	0.421	1
C3	0.125	1.28	1.155	1
C4	0.5	0.005	-0.495	0

Classify buyers of most-common title from a Korean E-Book retailer



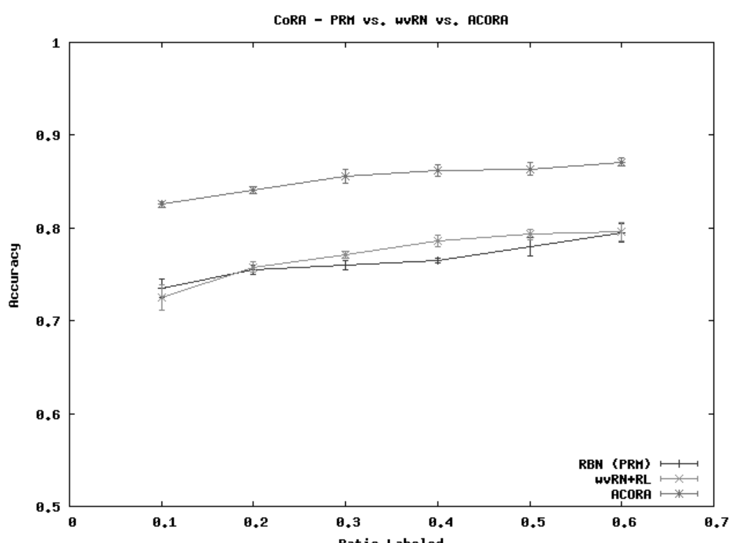
Class-conditional distributions across identifiers of 10 other popular books

Machine Learning Research Papers (from CoRA data)



(recall CoRA from discussion of univariate network models)

Using identifiers on CoRA



(Perlich & P. 2006)

(compare: Hill & P. “The Myth of the Double-Blind Review”, 2003)

Summary: Unique Characteristics of Networked Data (for predictive inference)

1. "Labeled" entities linked to "unlabeled" entities
   - allows "guilt-by-association" and related techniques
   - autocorrelation among neighbors
2. Collective inference is possible
   - inferences about entities can affect each other
3. Other aspects of neighbors can affect inferences about an entity
4. Identifiers can play an important role in modeling
   - being connected to specific individuals can be telling

➢ *Results show that there is a lot of power for prediction just in the network structure*

(3) Modeling for explanation using networked data

Using networked data to explain

Goals of this part of the tutorial

- Recognize the difference between the “simple” approach of associating network properties with outcomes and the emerging modern structural approaches that emphasize identification.
- Become familiar with a couple of examples of properties that have been useful in explaining ecommerce outcomes.
- Become familiar with a couple of emerging modern structural approaches to modeling networks that will lead to econometrically rigorous explanatory models.

Recall: Network properties

- Degree distribution
  - Extent of and variation in “local connectedness” across nodes
- PageRank
  - Extent of and variation in “centrality” across nodes
- Clustering
  - Extent of and variation in “shared connectedness” across nodes
- Average distance (diameter)
  - Extent of and variation in distance between nodes
- Assortative mixing/Homophily
  - Extent of and variation in “within-class connectedness” across nodes
- Distribution of components
- Degree correlation, community structure

The “simple” approach

- Theorize (perhaps using a mathematical model) how certain network properties will affect certain outcomes
  - Centrality and success
  - In-degree and income
  - Centrality and demand patterns
- Measure properties, outcomes
- Establish association between properties and outcomes by estimating reduced form equations.

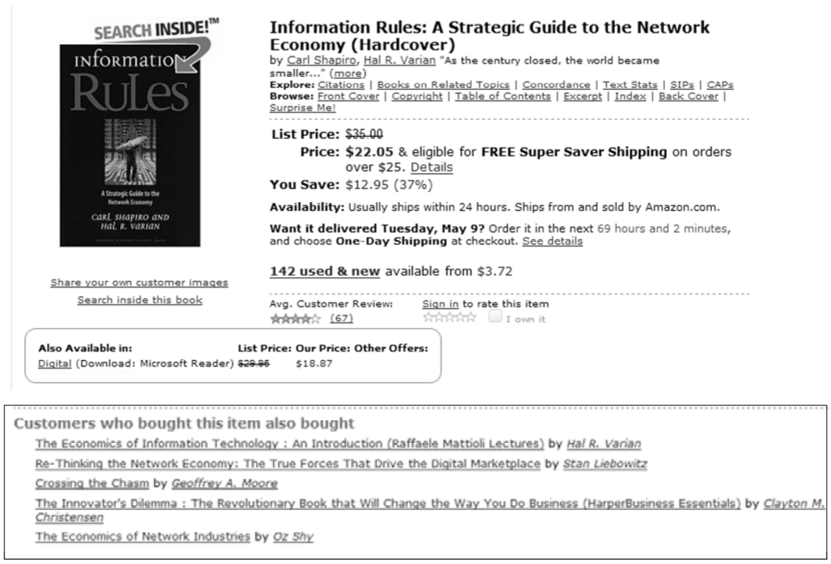
- Useful to establish co-movement, impossible to ascribe causation in a scientific way, widely used.

Example: PageRank and the long tail

- Degree distribution
  - Extent of and variation in “local connectedness” across nodes
- PageRank
  - Extent of and variation in “centrality” across nodes
  - Measure of “how important”, also “how influenced”

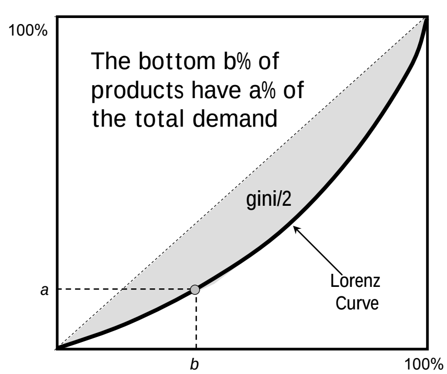
$$PageRank(i) = \frac{(1-\alpha)}{n} + \alpha \sum_{j \in G(i)} \left(\frac{PageRank(j)}{OutDegree(j)} \right)$$

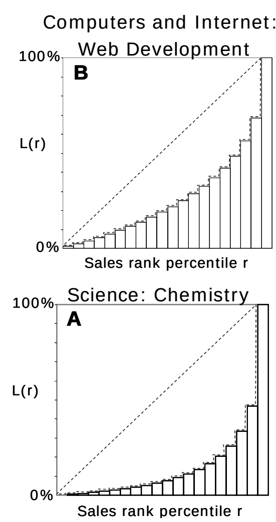
Example: PageRank and the long tail



Example: PageRank and the long tail

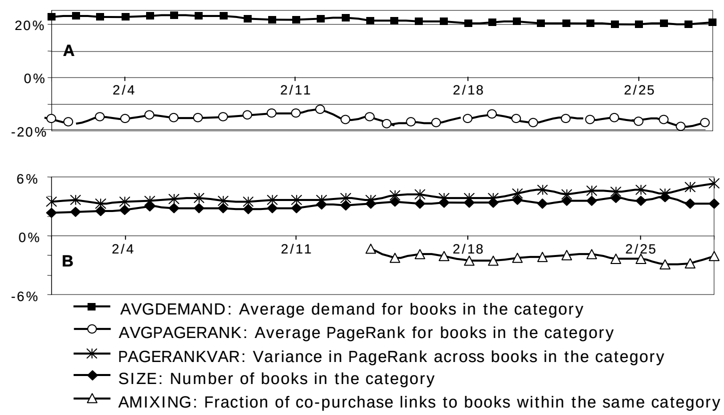
- Gini coefficient
  - Captures the extent to which demand is concentrated among the highest selling products in a group. Measured by the area above the Lorenz curve.





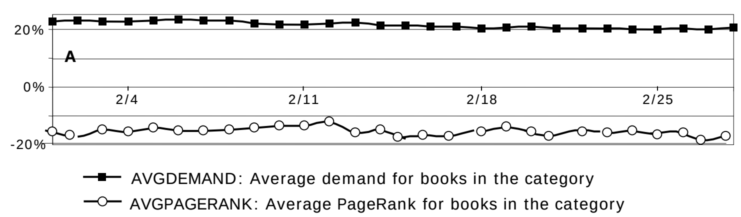
Example: PageRank and the long tail

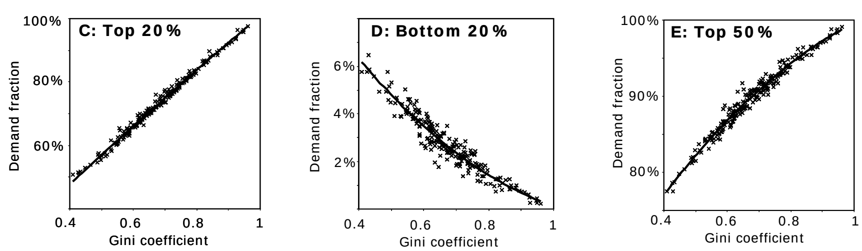
$$Log[\text{GINI}] = a + b_1 Log[\text{AVGDEMAND}] + b_2 Log[\text{AVGPAGERANK}] + b_3 Log[\text{PAGERANKVAR}] + b_4 Log[\text{SIZE}] + b_5 Log[\text{AMIXING}]$$



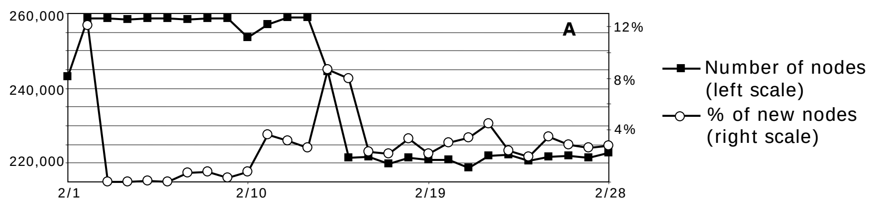
Example: PageRank and the long tail

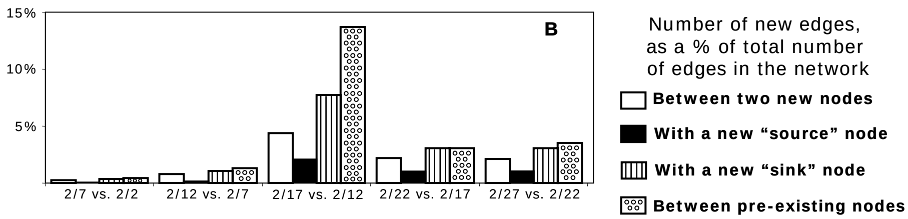
$$Log[\text{GINI}] = a + b_1 Log[\text{AVGDEMAND}] + b_2 Log[\text{AVGPAGERANK}] + b_3 Log[\text{PAGERANKVAR}] + b_4 Log[\text{SIZE}] + b_5 Log[\text{AMIXING}]$$





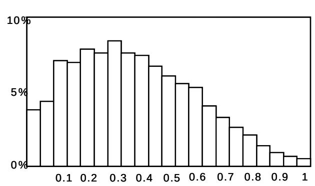
Other co-purchase graph properties



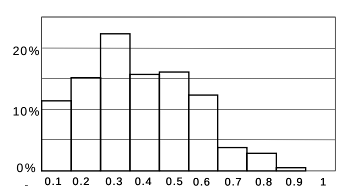


Other co-purchase graph properties

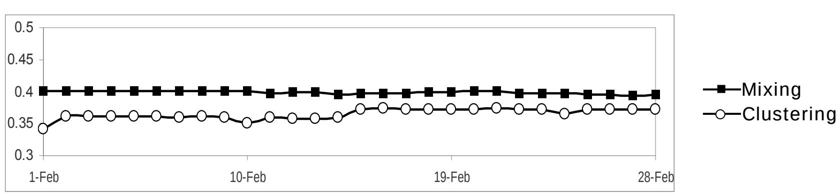
Random utility versus "location" model of choice?



Distribution of clustering coefficients for a sample day



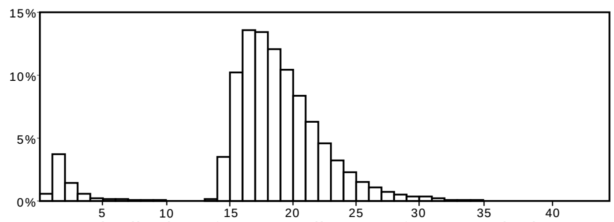
Distribution of average assortative mixing by category for a sample day



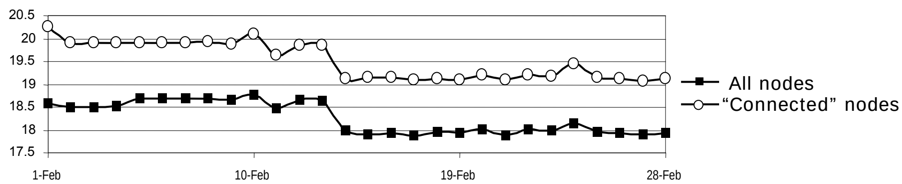
Average clustering coefficients, assortative mixing over a month

Other co-purchase graph properties

Nineteen degrees of separation?



Distribution of average distances for a sample day



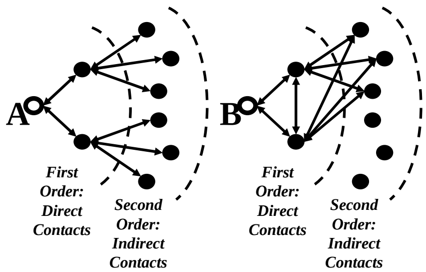
Average average distances over one month

Example: Position and Info. Advantage

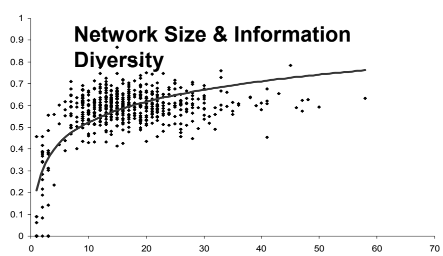
Aral and Van Alstyne (2007)

- The network: email communication between employees in an organization
- Establishes a relationship between network position and the diversity of information an employee has access to.
- Associates these two effects with employee productivity.
- Findings:
  - Larger, more diverse networks <-> more diverse information
  - More diverse information <-> Higher productivity
  - Diverse networks could play a role beyond simply providing more diverse information
- Communication trails will be increasingly common networked data sets in the future…

Example: Position and Info. Advantage



Network diversity



Dependent Variable:	Information Diversity	Information Diversity
Specification	*Fixed Effects*	*OLS-c*
Controls	*Age, Gender, Education, Industry Experience, Partner, Consultant*	
Total Email Incoming	-.001 (.001)	.001 (.001)
Network Size	.474*** (.114)	.296* (.138)
Network Size-Squared	-.272** (.089)	-.240* (.139)
Network Diversity	.128** (.052)	.268*** (.072)
Structural Equivalence	-.005 (.033)	.062 (.096)
Constant	.128* (.075)	.016 (.634)
Temporal Controls	Month	Month
F-Value (d.f.)	5.61*** (13)	5.03*** (19)
R^2	.14	.24
Obs.	540	434

Structural models to identify

- Identification (vastly simplified): recovering structural equation coefficients from reduced-form estimates.
- Identification in networked data is <u>hard</u>

- Background: Peer effects (Manski, 1993)

• y: outcome; x: characteristics; G: matrix defining "groups"

$$y = \alpha_0 + \alpha_1 Gx + \alpha_2 Gy + \alpha_3 x$$

Outcome vector | Exogenous (contextual) effect | Endogenous effect | Effect of own characteristics

• Real social effects cannot be separated from correlated effects
• The "reflection problem" makes identifying the endogenous effects from the exogenous effects hard.

Identifying peer effects

- Often in networked data, the “groups” associated with each observation are sufficiently different from each other.

$$y = \alpha_0 + \alpha_1 Gx + \alpha_2 Gy + \alpha_3 x$$

- Under certain linear independence properties of G, this facilitates the identification of social effects (Bramoulle et al, 2007)
- Other useful references: Lee (2003, 2006), Moffitt (2001)

Identifying peer effects

- The prior discussion helps identify social effects, but does not actually solve the problem of identifying the effects associated with the presence of an edge in a networked data set. (Notice that y is on both sides of the equation below, or there are contemporaneous and sometimes reciprocal effects of peers.)

$$y = \alpha_0 + \alpha_1 Gx + \alpha_2 Gy + \alpha_3 x$$

- One possible approach:
  - Estimate the outcome variables – that is, whatever you are trying to show is influenced by the edge – using only the exogenous variables (spatial autoregressive)
  - Compute estimates of the outcomes (endogenous variables) using these coefficients
  - Use these estimated endogenous variables in complete model
- Example: Peer effects and recommendation networks (Oestreicher-Singer and Sundararajan 2007).

Co-evolution of networks and behavior

- In some situations, networks influence behaviors (or outcomes), which in turn influence the networks over time. Recall the examples that this tutorial started with.
- A structural (and somewhat integrative) approach based on a more complete model of this kind is attempted by Snijders and coauthors (2004, 2005, 2007)

Co-evolution of networks and behavior

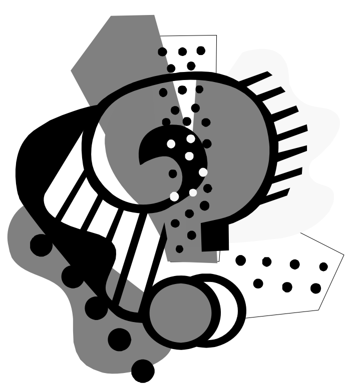
- Basic idea (and analogy with discrete choice logit):
  - Create a simplified (but internally complete) Markovian dynamic model of the co-evolution of networks and behaviors or actors in which all current outcomes and the current network are collectively the state, and all changes to outcomes and the network are “chosen” by actors.
  - Use one or a combination of a number of network properties to describe utility to each actor from each choice.
  - Estimate the parameters of this model directly (typically, maximum likelihood/Bayesian is not possible, and MCMC is required)

Explanation vs. Prediction

(intentionally blank)

Theories from the social sciences matter, whatever your research or business objective.

Questions and Discussion



http://pages.stern.nyu.edu/~fprovost/

http://oz.stern.nyu.edu/

Bibliography

J. C. Almack. The influence of intelligence on the selection of associates. School and Society, 16: 529–530, 1922.

Aral, S., Brynjolfsson, E. and M. Van Alstyne, 2006. Information, Technology & Information Worker Productivity: Task Level Evidence. International Conference on Information Systems 2006, Milwaukee, WI.

Aral, S. and M Van Alstyne, 2007. Network Position and Information Advantage. Working Paper, New York University.

Bala, V.,and S. Goyal, 2000. A Non-cooperative Model of Network Formation. *Econometrica* 68, 1181-1231.

A. Bernstein, S. Clearwater, and F. Provost. The relational vector-space model and industry classification. In Proceedings of the Learning Statistical Models from Relational Data Workshop at the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI), 2003.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society, 36(2):192–236, 1974.

J. Besag. Statistical analysis of non-lattice data. The Statistician, 24(3):179–195, 1975.

J. Besag. On the statistical analysis of dirty pictures. Journal of the Royal Statistical Society, 48(3): 259–302, 1986.

P. M. Blau. Inequality and Heterogeneity: A Primitive Theory of Social Structure. New York: Free Press, 1977.

A. Blum and S. Chawla. Learning from labeled and unlabeled data using graph mincuts. In Proceedings of the Eighteenth International Conference on Machine Learning (ICML), pages 19–26, 2001.

A. Blum, J. Lafferty, R. Reddy, and M. R. Rwebangira. Semi-supervised learning using randomized mincuts. In Proceedings of the Twenty-first International Conference on Machine Learning (ICML), pages 97–104, 2004.

H. Bott. Observation of play activities in a nursery school. Genetic Psychology Monographs, 4:44–88, 1928.

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 23(11):1222–1239, 2001.

Bramoulle, Y., and R. Kranton, 2005. Strategic Experimentation in Networks. *Journal of Economic Theory* (forthcoming).

Bramoulle, Y., Djebbari, H., and Fortin, B. (2007). Identification of peer effects through social networks. *Bonn Institute for the Study of Labor Working Paper*.

S. Chakrabarti, B. Dom, and P. Indyk. Enhanced hypertext categorization using hyperlinks. In Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data, pages 307–319, 1998.

Chwe, M., 2000. Communication and Coordination in Social Networks. *Review of Economic Studies* 67, 1-16.

C. Cortes, D. Pregibon, and C. T. Volinsky. Communities of interest. In Proceedings of the Fourth International Conference on Advances in Intelligent Data Analysis (IDA), pages 105–114, 2001.

R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. Probabilistic networks and expert systems. Springer, 1999.

M. Craven, D. Freitag, A. McCallum, T. Mitchell, K. Nigam, and C. Y. Quek. Learning to extract symbolic knowledge from the World Wide Web. In Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI), pages 509–516, 1998.

L. De Raedt, H. Blockeel, L. Dehaspe, and W. Van Laer. Three companions for data mining in first order logic. In S. Dzeroski and N. Lavrac, editors, Relational Data Mining, pages 105–139. Berlin; New York: Springer, 2001.

I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 269–274, 2001.

R. L. Dobrushin. The description of a random field by means of conditional probabilities and conditions of its regularity. Theory of Probability and its Applications, 13(2):197–224, 1968.

P. Domingos and M. Richardson. Mining the network value of customers. In Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 57–66, 2001.

S. Dzeroski and N. Lavrac. Relational Data Mining. Berlin; New York: Springer, 2001.

T. Fawcett and F. Provost. Adaptive fraud detection. Data Mining and Knowledge Discovery, 3: 291–316, 1997.

T. Fawcett and F. Provost. Activity monitoring: Noticing interesting changes in behavior. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 53–62, 1999.

P. A. Flach and N. Lachiche. Naive Bayesian classification of structured data. Machine Learning, 57:233-269, 2004.

L. Freeman. Some antecedents of social network analysis. *Connections* 19:39–42, 1996.

N. Friedman, L. Getoor, D. Koller, and A. Pfeffer. Learning probabilistic relational models. In Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI), pages 1300–1309, 1999.

Galeotti, A., S. Goyal, M. Jackson, and L. Yariv, 2006. Network Games. Mimeo, Caltech.

A. Galstyan and P. Cohen. Is guilt by association a bad thing? In Proceedings of the First International Conference on Intelligence Analysis (IA), 2005.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 6:721–741, 1984.

W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Markov Chain Monte Carlo in Practice. Chapman & Hall/CRC, 1995.

D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. Journal of the Royal Statistical Society, 51(2):271–279, 1989.

Goyal, S., and J. L. Moraga-Gonzalez, 2001. R&D Networks. The RAND Journal of Economics 32, 686-707.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer Verlag, New York, 2001.

D. D. Heckathorn and J. Jeffri. Jazz networks: Using respondent-driven sampling to study stratification in two jazz musician communities. Unpublished paper presented at American Sociological Association

Annual Meeting, August 2003.

D. Heckerman, D. M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie. Dependency networks for inference, collaborative filtering, and data visualization. Journal of Machine Learning Research (JMLR), 1:49–75, 2000.

S. Hill, F. Provost, and C. Volinsky. Network-based marketing: Identifying likely adopters via consumer networks. Statistical Science, 22(2):256–276, 2006a.

S. Hill, D.K. Agarwal, R. Bell, and C. Volinsky. Building an effective representation for dynamic networks. Journal of Computational & Graphical Statistics, 15(3):584–608, 2006b.

G. E. Hinton and T. J. Sejnowski. Learning and relearning in Boltzmann machines. Parallel Distributed Processing: Explorations in the Microstructure of Cognition, 1: Foundations:282–317, 1986.

J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences of the United States of America, 79(8): 2554–2558, 1982.

Z. Huang, H. Chen, and D. Zeng. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Transactions on Information Systems (TOIS), 22(1): 116–142, 2004.

R. A. Hummel and S. W. Zucker. On the foundations of relaxation labeling processes. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 5(3):267–287, 1983.

E. Ising. Beitrag zur Theorie des Ferromagnetismus. Zeitschrift f. Physik, 31:253–258, 1925. [German].

Jackson, M., 2004. A Survey of Models of Network Formation: Efficiency. Chapter 1 in Demange, G. and M. Wooders (eds.) Formation in Economics; Networks, Clubs and Coalitions, Cambridge Press.

Jackson, M., and L. Yariv, 2006. Diffusion of Behavior Properties in Network Games. Mimeo, Stanford University. http://www.stanford.edu/~jacksonm/netbehavior.pdf

M. Jackson, `The Study of Social Networks In Economics," forthcoming in The Missing Links: Formation and Decay of Economic Networks, edited by Joel Podolny and James E. Rauch; Russell Sage Foundation, 2007.

D. Jensen and J. Neville. Data mining in social networks. In National Academy of Sciences Symposium on Dynamic Social Network Modeling and Analysis, 2002a.

D. Jensen and J. Neville. Linkage and autocorrelation cause feature selection bias in relational learning. In Proceedings of the Nineteenth International Conference on Machine Learning (ICML), pages 259–266, 2002b.

D. Jensen, J. Neville, and B. Gallagher. Why collective inference improves relational classification. In Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 593–598, 2004.

T. Joachims. Transductive learning via spectral graph partitioning. In Proceedings of the Twentieth International Conference on Machine Learning (ICML), pages 290–297, 2003.

Kakade, S., M. Kearns and L. Ortiz, 2004. Graphical Economics. Available at http://www.cis.upenn.edu/~mkearns/papers/graphecon.pdf

Kelejian, H.H., and Prucha, I.R. (1998). A generalized spatial two-stage least square procedure for estimating a spatial autoregressive model with autoregressive disturbance. Journal of Real Estate Finance and Economics 104: 219-257.

K. Kersting, and L. De Raedt (2001). Towards combining inductive logic programming with Bayesian networks. In *Proceedings of the Eleventh International Conference on Inductive Logic Programming* (pp. 118–131). Strasbourg, France: Springer.

J. Kleinberg and ´ E. Tardos. Approximation algorithms for classification problems with pairwise relations: Metric labeling and Markov random fields. In Proceedings of the Fortieth Annual Symposium on Foundations of Computer Science (FOCS), pages 14–23, 1999.

R. Kohavi and G. John. Wrappers for feature subset selection. Artificial Intelligence, 97(1–2): 273–324, 1997.

D. Koller and A. Pfeffer. Probabilistic frame-based systems. In Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI), pages 580–587, 1998.

S. Kramer, N. Lavrac, and P. Flach. Propositionalization approaches to relational data mining. In S. Dzeroski and N. Lavrac, editors, Relational Data Mining, pages 262–291. Berlin; New York: Springer, 2001.

J. Lafferty, A. McCallum, and F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning (ICML), pages 282–289, 2001.

P. Langley. Crafting papers on machine learning. In Proceedings of the Seventeenth International Conference on Machine Learning (ICML), pages 1207–1212, 2000.

P. Lazarsfeld and R. K. Merton. Friendship as a social process: A substantive and methodological analysis. In M. Berger, T. Abel, and C. H. Page, editors, Freedom and Control in Modern Society, pages 18–66. Van Nostrand, 1954.

Lee, L.F. (2003). Best spatial two stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22: 307-335.

Lee, L.F. (2006). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Mimeo Ohio State University.

C. P. Loomis. Political and occupational cleavages in a Hanoverian village. Sociometry, 9:316–3333, 1946.

Q. Lu and L. Getoor. Link-based classification. In Proceedings of the Twentieth International Conference on Machine Learning (ICML), pages 496–503, 2003.

S. Macskassy and F. Provost. A simple relational classifier. In Proceedings of the Multi-Relational Data MiningWorkshop (MRDM) at the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003.

S. Macskassy and F. Provost. Suspicion scoring based on guilt-by-association, collective inference, and focused data access. In Proceedings of the First International Conference on Intelligence Analysis (IA), 2005.

S. Macskassy and F. Provost. Classification in Networked Data: A toolkit and a univariate case study. *Journal of Machine Learning Research* 8(May):935-983, 2007.

Manski, C.F. (1993). Identification of endogenous social effects: the reflection problem. Review of Economic Studies 60: 531-542.

A. McCallum, K. Nigam, J. Rennie, and K. Seymore. Automating the construction of internet portals with machine learning. Information Retrieval, 3(2):127–163, 2000.

M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. Annual Review of Sociology, 27:415–444, 2001.

Mitchell, T.M.. *Machine Learning*. McGraw Hill, New York, NY, 1997.

Mobius, M and A. Sdezl.. 2007. Trust and Social Collateral. http://mobius.fas.harvard.edu/papers/socialcollateral.pdf

Moffitt, R.A. (2001). Policy interventions, low-level equilibira, and social interactions, in S. Durlauf and H.P. Young, eds., Social Dynamics, MIT Press, Cambridge, MA: 45-82.

K. Murphy, Y. Weiss, and M. I. Jordan. Loopy belief-propagation for approximate inference: An empirical study. In Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence (UAI), pages 467–475, 1999.

J. Neville and D. Jensen. Iterative classification in relational data. In Proceedings of the Workshop on Learning Statistical Models from Relational Data at the Seventeenth National Conference on Artificial Intelligence (AAAI), pages 13–20, 2000.

J. Neville and D. Jensen. Collective classification with relational dependency networks. In Proceedings of the Multi-Relational Data Mining Workshop (MRDM) at the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003.

J. Neville and D. Jensen. Dependency networks for relational data. In Proceedings of the Fourth IEEE International Conference in Data Mining (ICDM), pages 170–177, 2004.

J. Neville and D. Jensen. Leveraging relational autocorrelation with latent group models. In Proceedings of the Fifth IEEE International Conference in Data Mining (ICDM), pages 322–329, 2005.

J. Neville and D. Jensen. Relational dependency networks. Journal of Machine Learning Research (JMLR), 8(Mar):653–692, 2007.

J. Neville, D. Jensen, L. Friedland, and M. Hay. Learning relational probability trees. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 625–630, 2003.

J. Neville, O. Simsek, D. Jensen, J. Komoroske, K. Palmer, and H. Goldberg. Using relational knowledge discovery to prevent securities fraud. In Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 449–458, 2005.

M. E. J. Newman. Mixing patterns in networks. Physical Review E, 67, 2003. 026126.

Oestreicher-Singer, G. and Sundararajan, A., 2006. Network Structure and the Long Tail of Electronic Commerce. Working Paper CeDER-07-03, New York University. (Earlier versions in Proceedings of the 2007 International Workshop and Conference on Network Science, 2006 International Conference on Information Systems).

Oestreicher-Singer, G. and Sundararajan, A., 2007. Recommendation Networks and Peer Effects in Electronic Commerce. Mimeo, New York University.

S. Pandit, D. Chau, S. Wang and C. Faloutsos. NetProbe: A Fast and Scalable System for Fraud Detection in Proceedings of WWW 2007, Banff, Alberta, Canada, May 8-12, 2007.

J. Pearl. Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, 1988.

C. Perlich. Citation-based document classification. In Workshop on Information Technology and Systems (WITS), 2003.

C. Perlich and F. Provost. Aggregation-based feature invention and relational concept classes. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 167–176, 2003.

C. Perlich and F. Provost. Distribution-based aggregation for relational learning with identifier attributes.

Machine Learning, 62(1/2):65–105, 2006.

C. Perlich, F. Provost, and J. Simonoff. Tree induction vs. logistic regression: A learning-curve analysis. Journal of Machine Learning Research (JMLR), 4:211-255, 2003.

A. Popescul and L. H. Ungar. Statistical relational learning for link prediction. In Proceedings of the Learning Statistical Models from Relational Data Workshop at the Nineteenth International Joint Conference on Artificial Intelligence (IJCAI), 2003.

R. B. Potts. Some generalized order-disorder transformations. Cambridge Philosophic Society, 48: 106–109, 1952.

Radner, R. and Sundararajan, A., 2006. Dynamic Pricing of Network Goods with Boundedly Rational Consumers. Working Paper CeDER-06-03 (2005 Winter Meeting of the Econometric Society).

H. M. Richardson. Community of values as a factor in friendships of college and adult women. Journal of Social Psychology, 11:303–312, 1940.

M. Richardson and P. Domingos. Markov logic networks. Machine Learning, 62(1/2):107–136, 2006.

J. Rocchio. Relevance feedback in information retrieval. In G. Salton, editor, The SMART Retrieval System: Experiments in Automatic Document Processing, chapter 14, pages 313–323. Prentice–Hall, 1971.

A. Rosenfeld, R. Hummel, and S. Zucker. Scene labeling by relaxation operations. IEEE Transactions on Systems, Man and Cybernetics, 6:420–433, 1976.

L. J. Savage. The Foundations of Statistics. John Wiley and Sons, 1954.

E. Segal, H. Wang, and D. Koller. Discovering molecular pathways from protein interaction and gene expression data. Bioinformatics, 19:I264–I272, Jul 2003a.

E. Segal, R. Yelensky, and D. Koller. Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. Bioinformatics, 19:I273–I282, Jul 2003b.

Tom A.B. Snijders, Christian E.G. Steglich, and Michael Schweinberger. Modeling the co-evolution of networks and behavior. Pages 41-71 in *Longitudinal models in the behavioral and related sciences*, edited by Kees van Montfort, Han Oud and Albert Satorra; Lawrence Erlbaum, 2007.

Steglich, C., Snijders, T.A.B., Pearson, M. Dynamic networks and behavior: Separating selection from influence. Mimeo University of Groningen, The Netherlands, 2004.

A. Sundararajan, 2004a. Nonlinear Pricing and Type-Dependent Network Effects. *Economics Letters* 83, 107-113.

A. Sundararajan, 2004b. Adoption Games in Networks: A Generalized Random Graph Model. Mimeo, New York University.

A. Sundararajan, 2007a. Local Network Effects and Complex Network Structure. *Contributions to Theoretical Economics* 7(1).

A. Sundararajan, 2007b. Network Seeding and Influence Diffusion. Mimeo, New York University.

B. Taskar, P. Abbeel, and D. Koller. Discriminative probabilistic models for relational data. In Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI), pages 485–492, 2002.

B. Taskar, E. Segal, and D. Koller. Probabilistic classification and clustering in relational data. In Proceedings of the Seventeenthth International Joint Conference on Artificial Intelligence (IJCAI), pages 870–878, 2001.

B. Taskar, V. Chatalbashev, and D. Koller. Learning associative Markov networks. In Proceedings of the Twenty-first International Conference on Machine Learning (ICML), 2004.

Tucker, C., 2006. Interactive, Option-Value and Domino Network Effects in Technology Adoption. *http://www.mit.edu/people/cetucker/videomessaginglocal.pdf*

K. Tumulty. Inside Bush's Secret Spy Net [electronic version]. Time, 167(21), 2006. Retrieved May 25, 2006, from http://www.time.com/time/archive/preview/0,10987,1194021,00.html.

V. N. Vapnik. The support vector method of function estimation. In J. Suykens and J. Vandewalle, editors, Nonlinear Modeling: Advanced Black-Box Techniques, pages 55–86. Kluwer, Boston, 1998a.

V. N. Vapnik. Statistical Learning Theory. John Wiley, NY, 1998b.

M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. Technical Report 649, University of California, Berkeley, 2003.

G. Winkler. Image Analysis, Random Fields and Markov Chain Monte Carlo Methods. Springer-Verlag, 2nd edition, 2003.

I. H. Witten and E. Frank. Data Mining: Practical Machine Learning Tools with Java Implementations. Morgan Kaufmann, San Francisco, 2000.

X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using Gaussian fields and harmonic functions. In Proceedings of the Twentieth International Conference on Machine Learning (ICML), pages 912–919, 2003.

Other Resources

Here is a non-exhaustive list of resources to explore work on complex networks, explanatory modeling (fairly thin) and predictive modeling with networked data (lots!). Beyond providing overviews and details, and identifying particular research projects, these resources give a flavor for the variety of topics, and a sampling of the researchers working on them.

- Books
  - Introduction to Statistical Relational Learning, ed. Getoor and Taskar 2007
    - http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=11331
  - Relational Data Mining, ed. Dzeroski and Lavrac 2001
    - http://www-ai.ijs.si/SasoDzeroski/RDMBook/
  - Random Graph Dynamics by Rick Durrett . Cambridge University Press, 2006
    - http://www.math.cornell.edu/~durrett/RGD/RGD.html
  - N.E.J Newman, The Structure and Function of Complex Networks. SIAM Review (this isn't a book but is better than any of the books that overview complex networks).
    - http://arxiv.org/abs/cond-mat/0303516
- Tutorial on Statistical Relational Learning
  - http://www.cs.umd.edu/~getoor/Talks/SRL-ICML-ILP05-Tutorial.ppt
- Tutorial on Complex Networks
  - http://cnls.lanl.gov/~ebn/cn/
- Resources for Social Network Analysis
  - http://stat.gamma.rug.nl/snijders/
- Special issues of the journal Machine Learning

- Multirelational data mining and statistical relational learning
    - http://www.springerlink.com/content/5830543713335321/
- Inductive logic programming
    - (several)
- Mining and Learning with Graphs
    - http://www.springer.com/cda/content/document/cda_downloaddocument/CFP_10994_171106.pdf?SGWID=0-0-45-334589-p35726603

- Conference on Social Networks
    - Sunbelt 2007: http://www.insna.org/2007/Sunbelt%202007.html
    - Sunbelt 2006: http://www.insna.org/2006/sunbelt2006.html
- Workshop on the Economics of Social Networks
    - ESSET 2006: http://www.szgerzensee.ch/research/conferences/esset06/?L=1
- Workshop on Statistical Network Analysis:
    - http://www.icml2006.org/icml2006/technical/workshops.html
- Workshops on statistical relational learning
    - ICML 2004 http://www.cs.umd.edu/projects/srl2004/
    - IJCAI 2003 http://kdl.cs.umass.edu/srl2003/
    - AAAI 2000 http://robotics.stanford.edu/srl
- Workshops on multi-relational data mining:
    - http://www-ai.ijs.si/SasoDzeroski/MRDM2004/
    - http://www-ai.ijs.si/SasoDzeroski/MRDM2003/
    - http://www-ai.ijs.si/SasoDzeroski/MRDM2002/
- Workshops on mining and learning with graphs
    - http://www.inf.uni-konstanz.de/mlg2006/index.shtml
    - http://mlg07.dsi.unifi.it/
    - (see also MGTS 2003-2005)
- Dagstuhl workshops on Probabilstic, Logical, & Relational Learning
    - http://www.dagstuhl.de/05051/
    - http://kathrin.dagstuhl.de/07161
- Conferences on Inductive Logic Programming (annual)
- NYU Workshops on the Economics of Information Technology
    - 2006: http://w4.stern.nyu.edu/ceder/events.cfm?doc_id=5583
    - 2005: http://w4.stern.nyu.edu/ceder/events.cfm?doc_id=4174